

16 July 1981

Nuclear issues in the Middle East

The conclusion that the Israeli raid on the Iraqi reactor at Tamuz on 7 June will have caused more diplomatic than physical damage is strengthened as the weeks pass. The organization now most on the defensive is not the government of Israel but the International Atomic Energy Agency in Vienna, which has had the invidious task in the past month of defending itself against the charge that its technical safeguards intended to prevent the diversion of fissile materials to military purposes are inadequate to the need. At the same time, many readers have taken offence at the leading article of 18 June which, among other things, argued that the Israeli attack was unjustifiable and that steps should now be taken, chiefly by the United States, to ensure that Israel signs the Non-Proliferation Treaty. The letters of protest have several themes in common — that the Tamuz raid was not a breach of international law because Iraq and Israel are technically in a state of war; that Iraq's motive for embarking on a nuclear programme cannot have been pacific, given its substantial reserves of petroleum; that Iraq's membership of the Non-Proliferation Treaty is irrelevant because the safeguards procedures are ineffective; and that in any case it is no business of a journal such as *Nature* to discuss political questions such as those prompted by the raid on Tamuz.

The real world, unfortunately, is neither as tidy nor as simple as these protests imply. Those who complain that a scientific journal has no place in politics are in general correct. It would, for example, be inappropriate (and deeply offensive to the bulk of readers) if a journal such as this were to take sides in a political election, "endorsing" this or that candidate for, say, the United States presidency (which is not, of course, to say that candidates' election promises about support for research are of no interest). There are, however, exceptions to the general rule. To the extent that complaints about the British government's policy on the universities chime in with those of the government's political opponents, they are inescapably political. And there is hardly any aspect of the Non-Proliferation Treaty, a matter of inescapable interest to the scientific community, that is strictly technical. The treaty, founded on the bargain between non-nuclear and nuclear states that the former would refrain from making weapons and that the latter would help with the development of civil nuclear technology, is intrinsically as much political as technical. More generally, it has been plain since the early 1960s that the widespread wish to inhibit the spread of nuclear weapons cannot be attained by strictly technical means; as the environmentalists used to say, there is no technological fix. Neither the treaty nor the safeguards elaborated since it came into force eleven years ago are a cast-iron assurance that weapons will not spread. International pressure and the fear of strategic countermeasures are essential ingredients in the now elaborate but untidy system of restraints that has developed. Those who consider that the reality is different should ask themselves why the first Indian nuclear explosion has not been followed by a second even after the passage of seven years.

But why should the scientific community have an interest in arrangements for inhibiting the spread of nuclear weapons? To quote J.R. Oppenheimer's remark at Alamogordo in 1946 that "the physicists have known guilt" would be to over-dramatize the case. Rather, the development of nuclear energy is the most striking illustration so far that research and development is rarely a benefit unalloyed. For the past three decades, the scientific community has been the chief source of policies for the sensible control of nuclear energy, both domestically and internationally.

Individual scientists have had a decisive influence on diplomatic innovations whose effect is to inhibit the spread of nuclear weapons. Thus the Partial Test-Ban Treaty is more a monument to Dr Jerome Wiesner's flair and persistence than to the good sense of his colleagues in the White House in the early 1960s. The safeguards system now operated by the International Atomic Energy Agency, although imperfect, is one of many illustrations that technical devices (and people) can serve wider interests. But to suppose that in such matters technical people should be inarticulate servants of the diplomats is not merely to ask too much of flesh and blood but to ignore the constructiveness of what the technical community has to offer. As it happens, the present is a time of exceptional anxiety about the future development and deployment of nuclear weapons. Can it seriously be suggested that technical people should help to work out and implement agreements of the kind for which people like Chancellor Helmut Schmidt have been asking for the past several months in strict innocence of the underlying issues? That may be a more substantial dereliction of responsibility.

So what are the issues provoked by the raid on Tamuz? The efficacy of the international agency's safeguards systems is certainly one of them. And it is proper openly to acknowledge that the safeguards system is not leak-proof. It never has been and never could be, which is why those who now pretend that the raid on Tamuz has for the first time revealed the truth are guilty of disingenuousness. For the system depends on the collaboration of national governments, which must disclose their activities in nuclear energy (and which could, in theory, keep some hidden). It supposes that inspections of, say, a reactor will reveal all, when some parts of all nuclear installations must be inaccessible on safety grounds. It relies to some extent for the detection of violations on the accumulation of information about a nuclear installation over a period of time, while on paper the treaty permits the resignation of any member state on three months' notice. It requires (but unnecessarily) that inspectors should keep the information they gather confidential, although the appearance of Dr Roger Richter before the Senate Foreign Relations Committee in Washington three weeks ago on the introduction of Senator Alan Cranston (who is not, as erroneously described last week, the chairman but only a member of the committee) shows that even that stipulation can be circumvented.

But so what? The Non-Proliferation Treaty is not the only treaty of its kind that relies on less than perfect arrangements for the detection of violations. Who, for example, would pretend that the requirements of the Threshold Test-Ban Treaty (which restricts the yield of underground nuclear tests to the equivalent of less than 150,000 tonnes of TNT but which relies on seismic measurements of actual yield that may be in error by a factor of two) are cast-iron? For at least two decades it has been accepted that no arms control treaty can embody cast-iron arrangements for verification. The objective is to ensure that the risk of detection is great enough to deter violation — and that the diplomatic risks of abrogation are equally forbidding. It was irresponsible of Senator Cranston to pretend that Dr Richter's evidence amounted to the first discovery of these truths, just as it was reprehensible of Dr Richter to give the Foreign Relations Committee a partial account of the internal discussions of the Iraqi reactor within the international agency. (Dr Hans Gruemm's account on page 187 is inherently more credible.)

Two other issues have been raised by correspondents, of which



the argument that the raid on Tamuz was legal because Israel and Iraq have never signed a peace treaty with each other is the least substantial. The delicacy of the military and political problems of the Middle East is notorious and several states, supporters of Israeli independence prominent among them, have spent much of their energy in recent years (no doubt partly from self-interest) seeking devices (of which Israel has known and approved) for making life less dangerous for Israelis and Iraqis. Mrs Jeanne Kirkpatrick, the United States representative at the United Nations, told the Security Council on 19 June that Israel had "hurt, not helped, the peace and security of the area". The same meeting of the council went on unanimously to adopt a resolution condemning the Israeli raid on Tamuz, asserting that Iraq should be compensated and demanding that Israel's nuclear facilities should be subject to international safeguards. Legality does not often excite such strictures.

What, then, happens next? The new government of France, which originally agreed to the contract under which the reactors at Tamuz were built, has fortunately reaffirmed its curious quasi-adherence to the Non-Proliferation Treaty — although not a signatory itself, it requires its nuclear customers to sign. Indeed, President Mitterrand, no doubt in reaction to recent bellicosity about nuclear weapons by President Hussein of Iraq, has made plain his intention of being even tougher in the future. In the United States, there is also now a prospect of a more rational anti-proliferation policy — the new Administration is apparently to announce this week its acceptance of the explicit bargain in the treaty that nuclear powers must be dependable sources of nuclear technology and materials if non-nuclear powers are to continue with their undertaking not to make nuclear weapons. (The promise does not distinguish between states with and without reserves of petroleum, nor should it.) There is much that could be done to improve the safeguards procedures of the international agency, but to the extent that non-nuclear powers may be required to assume obligations not originally contracted for, these questions should be raised at the next review conference of the treaty four years from now (which, as these things go, is almost tomorrow). Some attention should also be paid to recent Israeli advocacy of a nuclear-free zone in the Middle East — the stumbling block so far is that such a treaty would entail recognition of the state of Israel by its Arab neighbours and thus a resolution of the whole problem of the Middle East, but that is not a sufficient reason for ignoring the subject. The most immediate need, however, is that the nuclear powers and especially the United States should give whatever help they can to the international agency that has the invidious task of supervising the working of the Non-Proliferation Treaty. Behaving as if the treaty means what it says, which it does, is the most urgent need. Beyond that, it is essential that the nuclear powers signatory to the treaty should recall the reasons why they were given a drubbing at the second review conference in August last year, and that they should do what they can to implement those provisions of the Non-Proliferation Treaty that promise further steps (beyond Salt I) towards strategic arms control. Is it possible that Mr Eugene Rostow, now confirmed as director of the United States Arms Control and Disarmament Agency and a hawk in President Kennedy's administration, will be a dove in Mr Reagan's?

Lightning without cause

The British university system, more than ever alive with rumour when it is not already on vacation, is buzzing with a curious *canard*; is it possible, people are asking, that the way in which the cuts announced last week by the University Grants Committee have fallen has been surreptitiously guided, even invited, by stool-pigeons within the institutions concerned? This fanciful notion seems in many places to be the simplest explanation of how, without apparent consultation, the University Grants Committee has been uncannily perceptive in its suggestions of how individual universities should trim their sails. A university whose members have for years been complaining among themselves that too many

biological departments teach their own individual introductory courses is likely now to have been told to cut back on the teaching of "conventional biology". Chemistry departments here and there, whose members from time to time hang their heads with shame about the quality of their collective research, have found that their universities (against the general trend) have been advised to cut back on physical science. But who can have spilled the beans about this supposedly secret knowledge? Academics are understandably looking with suspicion at their vice-chancellors, those people who travel most often to London.

The truth is simpler. Supposedly secret knowledge is usually not as secret as its holders suppose. Many of the grants committee's negative decisions — "cut this, cut that" — are intelligible. For more than a year, for example, the committee has been urging that too many British universities have been making too much provision for the teaching of Russian; is it surprising that Russian departments were conspicuous among those marked out for oblivion two weeks ago? It is a misfortune that so few would-be students at British universities seek to be Russian experts. The grants committee's recommendations in other fields, and in the sciences especially, are harder to understand. Although some of the universities now complaining that they have been told to close departments superficially in the national interest know secretly that their case is only weak, others are genuinely (and rightly) puzzled. At least part of the explanation seems to be that the grants committee has relied on the public research councils for advice on the health and strength of research in departments of this or that at particular universities, discounting the importance of industrial support. But even more machiavellian calculations may have been involved; is it possible that the committee has assumed that the universities most seriously affected by its cuts are few enough in number to be able to capture support from industry? If so, the calculation is in error. Industrial companies will seek to back the winners in the committee's sweepstake, not the losers. And the chance of forcing a few British universities to make a new (and industrially-linked) way for themselves has been lost.

Science and technology in the new pattern of British universities will also be affected by a previously unremarked feature of the grants committee's general letter to universities — the absence (for the first time for many years) of "advice" about the way in which total student numbers should be partitioned between undergraduates and graduate students. Briefly, it will be left to universities individually to decide what balance should be struck. Those universities most seriously deprived of funds will decide, other things being equal as the economists say, to take in more undergraduates and fewer graduate students (who are more expensive). This simple calculation will, however, be muddled by calculations of the costs of dispensing with tenured faculty members. In the interests of the British university system, it would be a great advantage that postgraduate education should be more concentrated than it is, but it is unreasonable that universities that would benefit from following such a course should be expected to do so if the immediate consequence is that there is even less to spend on immediate needs — and no mechanism for making sure that bright undergraduates have a means of becoming graduate students themselves.

What this implies is that the grants committee's declarations about the immediate prospects for British universities cannot in themselves become a policy for the universities. They are at best a challenge to the other interested parties — the research councils, the universities themselves but also the government — to hammer out some kind of policy for the future. On present form, and that of the past two weeks, nothing much will happen. Most of those most closely concerned will be on vacation between now and when the cuts begin to bite. By then, it will be apparent that it is too late even for invigorated administrators of the British university system to exert much influence. The danger, as things have turned out, is that the academics most threatened by what the grants committee has decreed will be on holiday when crucial decisions must be made about the institutions for which they work. Too bad, people will be saying.

Row over nuclear safeguards continues

IAEA official denies evidence to Congress

The evidence given by Dr Roger Richter in the past few weeks to the foreign relations committees of the United States Congress was distorted, partly for effect and partly out of ignorance, according to Dr Hans Gruemm, the Austrian physicist who is head of the safeguards division of the International Atomic Energy Agency (IAEA). Dr Gruemm was speaking at a conference of the Institute of Nuclear Materials Handling in San Francisco on Monday this week.

In his address, Dr Gruemm acknowledged that the attack on the Iraqi reactor at Tamuz on 7 June, and the public comment that had followed — as well as Dr Richter's removal of confidential papers from the agency — had undermined the credibility of the safeguards system. He said, however, that the agency's internal memorandum dated 10 March, made public during the Senate hearings, was only one of several documents circulating within the agency and concerned with the improvement of safeguards on large research reactors.

The memorandum of 10 March, addressed to Dr Gruemm by Dr T. Shea of the agency's safeguards division, is a minute of a staff meeting on 12 February called to consider the feasibility of closing two loopholes in the agency's safeguards procedures — the possibility that a state covered by inspection might not have declared all the nuclear material used in its operations, and the possibility that reactors under safeguards might be used for making plutonium (or uranium-233) by the clandestine irradiation of natural uranium (or of thorium).

The meeting concluded that it would not be feasible to detect undeclared stocks of nuclear materials except by means of country-wide surveillance not covered by the safeguards provisions. On the use of reactors for the production of fissile material by neutron bombardment, the meeting apparently acknowledged that routine inspection would not always detect the use of uranium or thorium in this way because of the frequency with which samples for irradiation would normally be loaded and withdrawn (as in the production of short-lived isotopes), listed five technical devices that might serve continuously to monitor such illicit uses, agreed that none had been "identified" as "effective and efficient" and suggested further study. Dr Richter, in his account of these proceedings, did not say that the

agency was concerned.

Dr Gruemm said in San Francisco this week that the meeting in Vienna in February (at which Dr Richter had not been present) was only one of several technical meetings called to discuss improvements of safeguards procedures made possible by the recruitment of extra inspectors. He complained that Richter had not mentioned that studies of the feasibility of detecting the use of research reactors for manufacturing fissile material had been begun in 1980.

Dr Gruemm also said that detailed studies of the Tamuz reactor as a means of producing clandestine fissile material were begun at the end of 1979, when the agency was first informed of the transfer of natural and depleted uranium to Iraq. The agency's calculations showed that it might be possible to produce between one and two "significant quantities" (bomb units) of fissile material each year, but that this would be possible only by replenishing the

reactor core "several times a year" with enriched uranium, presumably from France. To make optimum clandestine use of the reactor, it would also have been necessary to add cooling circuits to the reactor, which would have been visible on inspection.

Dr Gruemm's statement in San Francisco is, however, unlikely soon to still the argument about the efficacy of the agency's safeguards, even though the agency appears to have resolved its previously chronic manpower problems — its professional safeguards staff increased fourfold (to 206) between 1970 and 1980.

Meanwhile, interest in the capacity of Israel to manufacture nuclear weapons has been revived by the report of a panel appointed by the United Nations, including Dr George Quester of Cornell University as its United States representative, that Israel is probably now in a position quickly to produce a substantial number of nuclear weapons.

Chevènement still battling for control

The *nouvelle politique* of science in France is off to a shaky start. Jean-Pierre Chevènement, Minister of State for Science and Technology, is still struggling with the Ministry of Industry for the control of certain key institutions; while on the other wing the more radical unions are calling for the resignation of certain research directors of the old regime.

The minister's contribution, announced last week, to the new 1981 budget — an adjustment to that of the previous government's — also falls a long way short of what he must aim for to bring French research and development spending up to the promised 2.5 per cent of gross national product by 1985. He will ask the National Assembly for an additional FF 154.9 million (around £15 million) compared with a total 1980 expenditure of FF 14,500 million, excluding defence research. That amounts to a 1 per cent increase; M. Chevènement will need ten times as much next year.

Of this year's extras, the minister will ask for FF 68.5 million to pay new salaries at the Centre National de la Recherche Scientifique (CNRS) — the main basic research agency — and at the Institut National de la Recherche Agronomique. Chevènement foresees the creation of 525 new posts, most of them for technicians and administrators. Then FF 25 million would prop up the contribution to the European centre for nuclear research (CERN) against the recent fall in the French franc. And FF 61.4 million would increase running budgets — but not of the basic research agencies. The Agence Nationale pour la Valorisation de la Recherche, which helps to convert discoveries in French laboratories into

innovations in French industry — and is rather a pet of M. Chevènement's — would get FF 60 million, and the remaining FF 1.4 million would go to the Centre National pour l'Exploitation des Océans to create fish farms.

Meanwhile Chevènement is struggling to establish his authority over the non-military research activities of the atomic energy and space authorities (CEA and CNES), and of the Agence Nationale pour la Valorisation de la Recherche itself, all of which at present belong to the Ministry for Industry. Chevènement's determination to control these agencies, in one way or another, and so to have — through the management of innovation — a major say in the socialist transformation of French industry, is not going down well among industry ministry bureaucrats, and they are now briefing their second minister in a few months, M. Pierre Dreyfus, to resist the scientific upstart.

However, Chevènement will almost certainly gain control of the Centre National de la Recherche Scientifique, where the unions — to whose interests Chevènement is sensitive — are becoming active. Most dramatically, the communist-affiliated Syndicat National des Chercheurs Scientifique last week called for the resignation of both the president and director of CNRS, Professor Charles Thibault and M. Jacques Ducuing. The director-general of the medical research agency, INSERM, M. Philippe Laudat, should also resign, said the union. They stand in the way of progress, according to the union secretary-general M. Michel Gruselle. Chevènement is cryptic; when asked recently if heads would roll, he said "we shall see".

Robert Walgate

Fusion research Where to next?

Brussels

The European Economic Community and its partner countries, Sweden and Switzerland, should go ahead with research into thermonuclear fusion — according to the European fusion review panel. The road to commercial fusion will, however, “be long and costly” and it appears unlikely that commercial fusion power will be in general use within the next 50 years. These are the main conclusions in the review panel’s report published last week.

The European Commission has also released the proposed programme for the next five years (1982–86) which will now be discussed in the committee of the member states’ permanent representatives (Coreper). It says that a major landmark in fusion research has almost been reached. There is a good prospect that the full objectives of JET (Joint European Torus) may be achieved and that the results could go beyond the stated objectives and attain thermonuclear ignition.

The construction of the basic JET device is now expected to be completed by the end of 1982 and the first discharge could take place in April 1983. A problem in going further, the review panel’s report points out, is that the device will become radioactive and hence inaccessible after experiments with tritium take place. But unless this step is taken, JET cannot be used to answer the crucial question of how the plasma behaves when alpha particles from the deuterium–tritium reaction are produced in large quantities. This knowledge is essential for progress towards the stage after JET.

This next step is called, with refreshing logic, NET (Next European Torus) and would demonstrate the technical feasibility of a commercial tokamak. The concept of an international commercial demonstration tokamak (INTOR) is being studied under the auspices of the International Atomic Energy Authority (IAEA). The review panel recommends that the Europeans need to make preparations for studies of the new technologies.

“But the Community will only be able to contribute to and benefit from INTOR if it develops in parallel its own design of a next step device and the technological

know-how necessary to undertake its construction. Both the Japanese and Americans are planning their own next step devices” warns the report.

The panel also weighs up the pros and cons of fusion research apart from the tokamaks and, not surprisingly at this early stage in the art, feels that options should be kept as wide open as possible. The three new specialized tokamak projects (TORE SUPRA, FTU, ASKEX-UPGRADE), which are at various stages of design, are also considered well justified.

Both the European Commission and the review panel favour the Reverse Field Experiment (RFX) proposed by laboratories in Culham, Padua and Los Alamos. An agreement has been drawn up under which the United States contributes \$8 million out of the total cost of \$48 million, but this awaits approval from the member states. The United States is also likely to participate in the two-stage development of the Advanced Stellarator. This would mean rebuilding the existing device in Garching, West Germany, and a new and much bigger stellarator later on.

All this research is likely to cost around 1,500 million European Units of Account (£740 million) over the next five years, according to the report: 400 million for JET, 560 million for the running costs of the associated laboratories, 125 million for investments in new supporting tokamaks, 260 million for the technology programme and NET and about 40 million for investments in the new alternative devices. This agrees with what the Commission has outlined.

The review panel’s report highlights several issues likely to give rise to political problems. One of these is the future of Culham. Should it be the site of the next step project and what should be done to prevent the tokamak becoming useless once it is radioactive? Also questioned is the belief that fusion is as unpolluting as has been argued — the radioactive waste will have to be stored. Finally, there is a danger that the duration of nuclear fusion research will outstrip the mortality of its acolytes. The average age of Community staff is 45 and the increase of average age is very nearly one year per year. Thus in about 15 years time when these staff retire, most of the know-how acquired in 30 to 40 years of research “will disappear rather suddenly.”

Jasper Becker

Clinch River reactor Start again folks

Washington

After four years of delay caused largely by President Carter’s reluctance to make a full-scale commitment to the “plutonium economy”, plans for the construction of a liquid metal fast breeder reactor at Clinch River in Tennessee have been put squarely back on the rails by the US Congress.

Despite support from President Reagan, it was never a foregone conclusion. In May, the Science and Technology Committee of the House of Representatives voted against funds for the initial construction work on the 350 megawatt reactor, which has been on the drawing board since the early 1970s, and for which over \$500 million worth of components have already been delivered.

In previous years the same committee had spearheaded congressional efforts which succeeded in preventing President Carter from killing the project; this time, cost-conscious Republicans added their voices to the previous minority which had argued that no further federal funds should be committed.

But larger political forces were at work. In particular, strong support for the construction of the Clinch River reactor came from Senate Leader Howard Baker, who represents Tennessee, and for whom the reactor means both jobs and votes. Reflecting Mr Baker’s wishes, the House accepted a Republican amendment to the budget bill at the end of last month which overturned the recommendations of the Science and Technology Committee and inserted \$230 million for preliminary construction of the reactor as proposed by Mr Reagan.

Similar approval has already been given by the Senate (which, ironically, has in past years voted to terminate the project). And although the decision to go ahead with construction must now pass through the appropriations process, it seems that after the crucial House vote this will be little more than a formality.

Opponents of the fast breeder maintain their opposition on several grounds. One is that the declining price of uranium over the past three years has reduced the need for a rapid breeder programme. The pressure has also been eased by declining predictions of future demand for electricity. And even many of those who support fast breeders in principle feel that the Clinch River design is now out-of-date.

The critics appear to have a covert ally in the new director of the Office of Management and Budget, Mr David Stockman. Long opposed to heavy federal subsidies for energy demonstration projects, Mr Stockman wrote to fellow Congressmen in 1977 stating explicitly that the federal government’s 90 per cent support for the Clinch River reactor was “totally incompatible with our free-market approach to energy policy”.



JET nears completion

Yet with over \$100 million of their own money already spent, the 733 utilities involved have been pushing for construction to go ahead. The level of jingoism has been high. Visitors to the Clinch River site are given a free coffee mug inscribed "we are fighting for energy independence". And in a letter expressing their support, 17 members of a group known as Scientists and Engineers for Secure Energy, headed by ex-president of the National Academy of Sciences, Frederick Seitz, gave as one of their reasons that "in view of recent political developments in certain Western countries, particularly France, the Clinch River Project may become the only reliable technological undertaking of its kind in the free world".

But in the end the personal support of Senator Howard Baker has been decisive. As characterized in the House budget bill, the liquid metal fast breeder reactor (LMFBR) will be one of a series of steps designed to bring US breeder technology in line with that of other industrialized nations.

The most recent of these steps has been the successful operation last December of the Fast Flux Test Facility (FFTF) at the Department of Energy's Hanford Reservation in Washington State. Late in March of this year, the 400 megawatt test facility emerged with flying colours from a safety test in which the reactor was shut down from full power, and the main coolant-circulating pumps were turned off. Construction of the Clinch River reactor, say its supporters, is the logical next step.

In approving the Clinch River funding (and therefore channelling support away from solar energy and conservation research which the Science and Technology Committee in the House of Representatives had preferred), the House authorized an initial \$15 million for the planning of a 1,000 megawatt reactor.

How much future support for the breeder programme will, in fact, be forthcoming from the Reagan Administration remains uncertain. In his formal presentations, budget director David Stockman has forsworn his earlier statements and repeated the Administration's support for LMFBR. In private, however, Mr Stockman and his officials at the Office of Management and Budget are said to be strongly opposed to further substantial government subsidies of the nuclear industry, including its fast breeder plans, preferring that the utilities should pay.

Meanwhile opponents have not given up the fight. They are giving wide publicity to the findings of a congressional investigation team that some of the contractors may have been overcharging for components already supplied. In the wings is a debate about whether the reactor meets the new siting requirements introduced by the Nuclear Regulatory Commission after the Three Mile Island accident. Congress may have made up its mind, but the public debate is far from over. **David Dickson**

Research ethics and safety

Changing the guard

Washington

In a small but symbolic way, last Thursday may turn out to be a significant turning-point in the history of public controls on genetic engineering. Meeting in Bethesda, Maryland, an advisory committee to the National Institutes of Health (NIH) decided to recommend to its parent body, the Recombinant DNA Advisory Committee (RAC), a further substantial relaxation of the safety controls applied to recombinant DNA research.

Meanwhile 50 miles away, in the depths of the Virginia countryside, a presidential commission established to look at the ethical problems raised by biomedical advances has suggested the establishment of a new body — possibly at an international level — charged to seek a social consensus on the various dilemmas which the expanding clinical use of genetic engineering techniques will raise.

The RAC subcommittee was set up at a meeting of the full committee in May to discuss proposals for a significant relaxation in the safety guidelines made by Dr David Baltimore of the Massachusetts Institute of Technology and Dr Allan Campbell of Stanford University (*Nature* 7 May, p.3).

The subcommittee, whose recommendations will now be discussed at the next full meeting of RAC in September, did not agree that NIH guidelines should be made voluntary. However, they did suggest that detailed rules for the composition of local institutional biohazards committees (IBC) be removed.

If eventually approved by NIH, this would mean that research institutions would no longer be required to include "public interest" representatives, for example, on their IBC (although many would probably continue to do so). It could also mean that the responsibilities of the IBC to ensure compliance with the guidelines could be delegated to a single institutional biosafety officer.

The subcommittee also supported a proposal to eliminate from the guidelines a detailed listing of containment procedures, and its replacement by a statement that such procedures should follow recommendations being developed by the Center for Disease Control for experiments using the host or the vector separately.

In other cases the subcommittee proposed that P1 containment levels be used, and that a statement be included about donor DNA, saying that if there is clear evidence that the donor DNA will significantly alter the pathogenicity of the host, then the appropriate containment level will be applied.

Some of the suggestions approved by the subcommittee — for example that all prohibitions requiring special permission from the director of NIH, including work with

Ziman speaks out

Professor John Ziman on Monday strongly criticized the Royal Society, of which he is a fellow, for sluggishness on human rights issues. He was addressing the All-Party Parliamentary Committee for Society Jewry at a special award ceremony in the House of Commons, at which he received on behalf of Dr Viktor Brailovskii, who last month was sentenced to five years' Siberian exile, a Henry Moore lithograph entitled "for courage in defence of freedom".

Professor Ziman earlier this year had received, in conjunction with Dr John Humphrey (until recently deputy director of the National Institute of Medical Research) and lawyer Paul Sieghart, the second annual Airey Neave award, which will finance a study of freedom in science. He was therefore an obvious proxy for Dr Brailovskii who, until his arrest last November, had acted as host and organizer of the Sunday seminar for Jewish "refusnik" scientists denied emigration visas but dismissed from their academic posts after applying for them.

Professor Ziman said it would have been more appropriate that Dr Brailovskii's proxy should have been not a private scientist such as himself but the president of the Royal Society in his official capacity. He recalled that Dr Aleksandr Voronel', the founder of the Sunday seminar, said in Britain shortly after being allowed to emigrate in 1974 that "the seminar is the only true representative of free and independent science in the Soviet Union". The Royal Society, whose official aim is "improving human knowledge", should therefore, said Professor Ziman, give the fullest possible support to the seminar — support which, so far, it has been reluctant to afford.

Vera Rich

cultures over 10 litres, be eliminated from the guidelines — went further than Dr Baltimore and Dr Campbell had proposed to RAC. Others, such as the continuation of local biosafety procedures and the recommendation that the guidelines remain mandatory for scientists working with NIH funds, are more conservative.

Any decisions by RAC at its next meeting will be subsequently published in the *Federal Register* for public comment. After that, the matter will rest with Dr Richard M. Krause, director of the National Institute of Allergy and Infectious Diseases, who was given full responsibility for ensuring compliance with the guidelines by Dr Donald Fredrickson when he resigned as NIH director on 30 June.

As the safety debate is being wound down at NIH, a complementary discussion about the steps necessary to prevent undesirable clinical applications of genetic manipulation techniques has been getting

under way with the President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research.

The commission was asked last year to look at the ethical issues raised by genetic engineering by President Carter's science adviser, Dr Frank Press, following a letter to the President from three church groups expressing concern that recent advances in genetic research meant that "those who would play God will be tempted as never before". Since then the churches involved have not demonstrated a particularly close interest, but public concern has been stimulated by various press reports of potential new surgical techniques.

At last week's meeting the members discussed a draft report on the ethical and social aspects of genetic engineering. And although reluctant to raise unnecessary fears, they agreed that the implications were likely to be significant — for example in terms of the potential ability of an individual to alter the genetic characteristics of his or her descendants.

Most commission members agreed that there was a need for a wider public dissemination of information about the potential effects of new clinical techniques. Also that it might be appropriate for some type of advisory body to indicate areas in which caution should be used.

There was less of a consensus on whether it was desirable that such a body should suggest that certain types of experiments — for example the cloning of a human being — should be prohibited. Some, for example, suggested that any attempt at what the draft report described as the "control of evolution" should be proscribed; others pointed out the phrase was so broad as to include many currently-accepted practices, such as the treatment of diabetes with insulin.

The commission also debated whether discussions should take place at an international level. There was general agreement, however, that achieving international consensus on the boundary between acceptable and unacceptable practices would be even more difficult than at a national level.

David Dickson

New substance regulations

Industry complains

The British Chemical Industries Association is protesting vigorously at the draft on the notification of new chemical substances drawn up by the Health and Safety Executive. The association claims that the draft regulations would mean the end of research and development in the British chemical industry.

The draft regulations were published last February, when comments from interested parties were invited by the end of this month. They are an attempt to bring British practice into line with a directive of the European Commission, whose aim is to

protect "man and the environment" from the potential hazards of new substances. Although the directive deals chiefly with the protection of the consumer, the Health and Safety Executive is (given its remit) primarily concerned with the protection of workers' health.

Thus the British regulations would require industry to notify not only all new manufactured substances but also all new intermediates in chemical processes. The Chemical Industries Association complains that the extra costs involved will drive research and development away from Britain. The dilemma is, however, real. The Health and Safety Executive says that intermediates must be tested if existing regulations to protect workers from potential hazards are to be put on a more formal basis than required by the Health and Safety at Work Act.

The consultative document is precise about the tests required for new chemicals. Manufacturers would be required to assess the toxicity of substances by LD₅₀ tests, provide data on skin and eye irritability, tests for mutagenicity and possibly carcinogenicity and teratogenicity. They would also have to provide data on biodegradability. The executive estimates that the total cost would be about £45,000 per substance.

The objections of the chemical industry appear to centre more on the range of chemicals covered than the direct cost. As well as chemical intermediates, the draft regulations cover pharmaceuticals, food-stuffs and pesticides, all of which are excluded from the European directive on the grounds that they are covered by other regulations. The Health and Safety Executive's argument is that such assessments relate only to specific uses, and are not necessarily sufficient.

After the July deadline, the chemical industry will also be arguing for a strengthening of the provisions for preserving confidentiality. The association is concerned that valuable data, especially those on intermediates which would provide information on novel process routes, could fall into competitors' hands.

So far, few other bodies have put in comments, but the trade unions and environmentalists will also be having their say. The controversy aroused by the chemical industry's response, however, suggests that further negotiation will be needed before the regulations are in a final form and that the European Commission's 18 September deadline for the implementation of legislation will not now be met.

The European Commission is at the same time going ahead with its plan to compile a catalogue of all chemicals manufactured in Europe. Thereafter, industry will be required to notify the commission of all new chemicals manufactured during the past ten years that are not included, a ruling that will apply even in countries that will not have introduced their own legislation.

Judy Redfearn

UK pharmaceutical industry

Keeping up

"Britain's medicine makers have brought out an 'unfashionable' annual report — it tells a success story." So says the cheery publicity blurb announcing the 1980–81 report of the Association of the British Pharmaceutical Industry (ABPI). The claimed success is an increase in the value of exports of pharmaceuticals to £745.4 million in 1980, 16.7 per cent up on the sales in 1979, and representing a £523 million surplus of exports over imports.

Evidence of success is rare enough in British industry and the bouncy confidence affected by ABPI is likely to please a government eager for good news. And the industry seems to be getting its views across with some aplomb. Already this year the industry has had a victory in the form of new regulations governing the granting of the clinical trial certificates which must be obtained before new drugs can be tested clinically. From March, the certification process requires simpler documentation and data requirements have been reduced.

In another area of concern to ABPI, Mrs Sally Oppenheim, Minister of State for Consumer Affairs, has said that the government will try to include a "state of the art" defence in the EEC directive on product liability. The objective is that manufacturers should not be held liable for injuries to health caused by a product which could not be termed "defective" in the light of scientific knowledge when the drug was put onto the market.

The supposed main benefits to be gained by the simpler clinical certification rules are a reduction in the 10 to 12 years now needed for a new drug to reach the patient (a debatable improvement, especially after Fisons' withdrawal of Proxicromil when it was all but on the market, see *Nature* 12 March, p.81), and a stimulus to encourage development of drugs to treat less common diseases.

However, the most tangible effect of relaxations in control of drugs in clinical trials is likely to be an increase in the numbers of trials conducted in the United Kingdom rather than in other, less restrictive, countries. This is one factor to be considered by multinational companies when deciding whether to invest in Britain or go elsewhere. At present investment in research in the United Kingdom is holding up well, with £16 million to be invested by Merck, Sharp and Dohme in a neurobiology research centre near Harlow, £5 million by Upjohn in expanding facilities at West Crawley, £10 million by the Wellcome Foundation in a chemical research laboratory in Beckenham, and £3.3 million by Roche in improving research facilities at Welwyn Garden City. But competition between the developed countries for the favours of the research-based companies can only increase.

Charles Wenz

Industry and research

Dupont joins in*Boston*

Reflecting the growing eagerness of industry and universities to find ways in which both can profit from advances in biotechnology, the chemical company Dupont has promised \$6 million for genetic engineering research to a group of workers at Massachusetts General Hospital, one of the teaching hospitals associated with Harvard University. This announcement comes just as scientists, politicians and university administrators are completing their examination of a ten-year, \$50 million grant to another group of Harvard geneticists proposed by the German pharmaceutical house Hoechst-Roussel (see *Nature* 18 June, p. 525).

The Dupont grant will support the new genetics department at the Harvard Medical School under Dr Philip Leder who is moving to Harvard in September from the National Institutes of Health. The conditions of the Dupont deal are similar to the Hoechst arrangement; the grant recipients retain patent rights on discoveries made with corporate funds, but the companies are given exclusive licences to develop any products that result. One difference in the Dupont grant is that the company has not requested training positions for its scientists.

Although most university officials applaud these agreements, in the past month committees in both the Senate and the House of Representatives, officials of the National Institutes of Health and high ranking university officials from the United States and Canada have met independently to review Harvard's new affiliations with industry and to question whether the quality and independence of research will be compromised by corporate support.

These hearings have attracted wide attention since Senator William Hatch complained that the Harvard-Hoechst grant fostered a "technology leak" by giving a German company first crack at exploiting discoveries at institutions supported by public funds. The victims of such a situation, he suggested, would be the US taxpayers who have indirectly supported basic research that will benefit foreign investors.

This concern will not be a problem with the Dupont grant, because the money will be coming from a US chemical and drug conglomerate. Dupont plans to give Harvard the \$6 million in annual instalments, with \$2 million immediately and \$1 million over each of the next four years.

University officials say that they will examine all patentable discoveries very carefully to determine whether they were made with federal funds or portions of the Dupont grant. Only the latter will bring Dupont exclusive licensing rights to

university-owned patents.

The toughest scrutiny, however, may come from the local city councils. The Cambridge City Council has officially ordered an investigation of Harvard gene-splicing experiments in tax-exempt laboratories. Council members intend to rescind tax exemptions where private corporations will reap financial benefits.

Harvard officials do not anticipate an investigation of the Dupont grant similar to the House investigation of the Massachusetts General Hospital-Hoechst agreement. **Michael D. Stein**

Chinese Academy of Sciences**Scientist takes over**

China's Academy of Sciences once more has a scientist at its head. A few weeks ago the academy's Scientific Council met for the first time for 21 years and elected as its president Dr Lu Jiaxi, a physicist, who at one time worked at University College London under Professor S. Sugden and at the California Institute of Technology under Linus Pauling.

Dr Lu takes over the presidency from Fang Yi, the minister responsible for the State Commission for Science and Technology. On resigning, Minister Fang stressed that the presidency should be an elected post (he had been a government appointee) and should go to a scientist.

At the same meeting, the Scientific Council accepted a draft constitution which states that the supreme decision-making power for the academy will be vested in the 400-strong council, which from now on should meet every two years. Between council meetings, the academy will be administered by a presidium of 30 members, of whom 20 will be members of the council and 10 from the Communist party and government departments.

These changes reflect an overall policy of increasing the professionalism and prestige of Chinese science, which suffered considerably during the Cultural Revolution of 1966-1976. As well as revitalizing the Academy of Sciences, recent developments include the conversion of a teacher's training college in Hainan into a university, which will commence undergraduate enrolment in autumn 1983 and major discussions on how to develop vocational and technical secondary education. This latter need is so pressing that one government spokesman suggested that in addition to the state system, private projects for training technical workers should be encouraged. And to help young people develop their interest in science, a new "National Association of Science Coaches for Juveniles" was established in Peking, with Jiang Nanxiang, the Minister of Education as Honorary Chairman, and Wu Zhonghua, executive chairman of the Presidium of the Academy of Sciences as chairman of the Board of Directors.

Vera Rich**Spanish research****Painful reform***Barcelona*

The Spanish Higher Research Council (Consejo Superior de Investigaciones Científicas (CSIC) is nearing the end of a process aimed at giving it a fresh start. The reformers, the present directorate of CSIC, intend to demonstrate to public opinion and to the government that CSIC can renew itself and become an efficient instrument in the economic and cultural development of Spain.

CSIC is the main public body devoted to research in Spain. It has a permanent staff of 5,000 (1,500 of them scientists) in about a hundred institutes. Since its foundation and haphazard development during the government of General Franco, CSIC has maintained an enormous diversity of research interests of variable quality. There are groups devoted to local studies and to theology, to textile technology and fisheries and also to history, biology and physics. Only a few of these institutes can be compared with advanced European or American laboratories, and most of them suffer from limited and ageing staff and poor funding. This has provoked criticism from both inside and outside CSIC, directed chiefly at the low scientific level of many institutes and their lack of relevance to the general needs of the country. During the past five years of democratic government, even the abolition of CSIC has been openly discussed.

A move to reverse this trend from inside CSIC began after the appointment of a new directorate last year. CSIC's Scientific Commission, consisting of scientists from different areas of research, was asked to suggest important areas of research and to advise laboratories how they might best pursue them. A number of commissions are now working on detailed programmes based on these suggestions. As CSIC's budget cannot finance all the programmes chosen, it is hoped that outside agencies will be able to provide additional funds. It is hoped that groups will join forces to produce plans that are likely to be financed.

CSIC scientists are anxious to give the proposed changes a chance of success, but there is some doubt about the speed at which the scheme is being carried out, particularly in the absence of any long-term decision at government level.

Increases of staff are unlikely because of cuts in public expenditure and there is doubt as to whether the proposed Law of University, including to provide guidelines for recruiting teaching and research staff, will be passed before the next elections in 1983. In the meantime, the need of people and the low level of salaries are increasing tension inside CSIC and could lead to the resignation of the present directorate. A welcome reform lies on a knife-edge.

Pedro Puigdoménech

CORRESPONDENCE

Unfair on DeVita

SIR — The recent US Senate hearings in which Dr Vincent DeVita was scourged for the failure of the NIH-NCI bureaucracy to vitiate promptly an accused laboratory cheat leaves one appalled and apprehensive. Dr DeVita is the first NCI director who is a clinical oncologist, an occupation faced with the unenviable repetitive experience of patients who become progressively worse, mortify and die. In no small part due to the efforts of Dr DeVita and the organization he has assembled, this no longer takes the invariable course it once did and the concept of a cure has become realistic for many cancer patients. He has tried to redirect the cancer research industry back to the human cancer patient and away from high-visibility science for science's sake, while still juggling the all important basic research that must be done as well. To hear of his being chastized in such a silly way can be best compared to rejecting Socratic reasoning because Socrates may have been bald.

Cecil H. Fox

Silver Spring,
Maryland, USA

Serbian exchange

SIR — We would like to report on the great value of a recent exchange visit between British and Yugoslav geologists. In May we took part in a ten-day field trip to Yugoslavia, organized jointly as part of an agreement between the Royal Society of London and the Council of Academies of Science and Arts of the Federative Socialist Republic of Yugoslavia. It followed a visit in autumn 1980 when Yugoslav geologists Professor V. Majer and S. Karamata came to the UK, the first time that such an exchange had occurred between geologists of these countries.

Our time was spent in Yugoslavia mostly on a geological excursion through Serbia by minibus organized by Professor S. Karamata of the Faculty of Mining and Geology at Belgrade University. A round trip of over 2,100 km was made, starting from Belgrade and passing through the Carpatho-Balkan chain, the Serbo-Macedonian massif and the Inner and eastern Outer Dinarides. Stops were made at selected localities of interest enabling the international team of British, French, Hungarian and Yugoslav geologists to exchange ideas and collect rock samples. The lithology observed ranged from Palaeozoic basement, Mesozoic lavas and sediments and Cenozoic calc-alkaline flows to ultramafic, mafic and metamorphic components of ophiolites and related olistostromes.

This traverse enabled us to obtain an excellent introduction to the geology of Yugoslavia — an essential part of studies in the Earth sciences where it is necessary to observe at first hand natural phenomena in the field. We hope this initial exchange will lead to further visits and joint research in the near future.

S. O. AGRELL
A. G. SMITH
J. G. SPRAY

Department of Earth Sciences,
University of Cambridge, UK

Human protection

SIR — An imprecise reference in a recent article, "NIH censure for Dr Martin Cline," (*Nature* 4 June, p.269) prompts us to delineate the distinction between the UCLA School of Medicine Human Subject Protection Committee which reviews research protocols involving human subjects for risk-benefit ratio and the UCLA Human Subject Policy Committee which, based on federal regulations, develops university policy governing research involving human subjects.

ESTHER F. HAYS

Human Subject Protection Committee,
School of Medicine,
University of California, Los Angeles, USA

Fight the obscure

SIR — Recognizing the present popularity of various forms of anti-scientific obscurantism, both here and still more in his own country, Dr Ralph Lewis was dead right to conclude his letter (*Nature* 11 June, p.448): "We must find and state clearly our fundamental agreements, so that our differences in presentation cannot be misconstrued and distorted".

I suggest that one first necessary step is to insist always that the contrary of any evolutionary account of the origin of species is a doctrine of special creation. Evolution in this understanding would be decisively refuted were palaeontologists to discover fossil remains of creatures morphologically indistinguishable from the higher mammals in rocks which are much too old — J.B.S. Haldane used to specify human remains in a coal seam. Darwinian and any alternative accounts of the mechanism of such evolution, or of its mechanisms — the Lamarckian inheritance of acquired characteristics, for instance — must all be equally inconsistent with any doctrine of special creation. And, although I will not myself pretend to understand cladistics, we can be sure that its Marxist adherents are not rooting for special creation as opposed to evolution. Their point against Darwin is, surely, that while he defended "that old canon in natural history, *Natura non facit saltum*", they themselves are committed to the opposite contention, that both Nature and humanity sometimes take "A great leap forward".

A second step is to dispose of certain philosophical muddles and misconceptions — such as those propagated by the British Museum (Natural History). Barry Cox quoted several in his report (*Nature* 4 June, p.373). One is that the assertion of the survival of the fittest is vacuously tautological: "The Survival of the Fittest is an empty phrase, it is a play upon words". Certainly, since the criterion of fitness to survive is here actual survival, this assertion is not to be construed as an assurance that all is for the best. It is, nevertheless, not a tautology. For clearly it denies that the survivors survive at random, while asserting that they have some sort of competitive edge over the non-survivors.

Again, Cox quotes the statement: "The idea of evolution by natural selection is a matter of logic, not science, and it follows that the concept of evolution by natural selection is not, strictly speaking, scientific". Certainly it is possible formally to deduce the conclusion

that some natural selection must occur from two or three general propositions about multiplicative reproduction, variation, and the finitude of resources. But this neither disqualifies that conclusion as science nor turns it into an empty truth of logic. Substantial truths can be and are deduced with perfect formal validity from premises stating other similarly substantial truths.

ANTONY FLEW

Department of Philosophy,
University of Reading, UK

Defining a lectin

SIR — The Nomenclature Committee of the International Union of Biochemistry has discussed the objections raised by Kocourek and Hořejší¹ to the definition of lectins proposed by Goldstein *et al.*² which was also published by the Nomenclature Committee in its 1981 Newsletter³.

According to both definitions, "a lectin is a sugar-binding protein of non-immune origin" (glycoproteins are a class of proteins and there is no need to specify them in the definition). The emphasis in the definition of Goldstein *et al.* is an operational one, so the definition is easy to apply experimentally by the criteria of agglutination of cells (not necessarily of erythrocytes, which carry a limited variety of sugars on their surface) or precipitation of glycoconjugates. By contrast, the criterion of Kocourek and Hořejší of lack of enzymatic activity is difficult to apply; for example, glycosyltransferase activity might be present with an acceptor different from those tested. Moreover, there is now evidence that certain proteins hitherto known as lectins possess glycosidase activity⁴.

The definition of Goldstein *et al.* implies the presence of more than one carbohydrate-binding site. Strictly speaking, it excludes proteins such as ricin that are closely related to lectins in certain properties (for example, acid composition and possibly primary structure) and derived from the same plants, and this may be a disadvantage. Nevertheless the definition of Kocourek and Hořejší is so close to meaning "carbohydrate-binding protein" that a special name seems unnecessary, and the definition is too broad to be useful, since it includes substances such as sugar-transport proteins, chemotaxis receptors, certain bacterial toxins, hormones and interferons.

As long as we do not know with certainty the role of lectins, whether in plants, animals or microorganisms, it seems preferable to focus the definition of these substances on positive and easily testable properties. We therefore plan to continue to use the word as proposed by Goldstein *et al.*

H.B.F. DIXON
(Secretary)

Nomenclature Committee of International
Union of Biochemistry,
Cambridge, UK

1. Kocourek, J. & Hořejší, V. *Nature, News & Views* **290**, 188 (1981).
2. Goldstein, I.J., Hughes, R.C., Monsigny, M., Osawa, T. & Sharon, N. *Nature* **285**, 66 (1980).
3. Nomenclature Committee *Archs Biochem. Biophys.* **206**, 458-462 (1981); *Eur. J. Biochem.* **114**, 1-4 (1981); *Hoppe-Seyler's Z. physiol. Chem.* **362**, I-IV (1981); *J. biol. Chem.* **256**, 12-14 (1981).
4. Hankins, C.N. & Shannon, L.M. *J. biol. Chem.* **253**, 7791-7797 (1978); *Pl. Physiol.* **65**, 618-622 (1980).

NEWS AND VIEWS

Hydrothermal processes on ridge flanks: the *Challenger's* return to the Costa Rica rift

from Roger N. Anderson

THIS coming November the drilling vessel *Glomar Challenger* will reoccupy a site which has already produced some of the most spectacular scientific discoveries of the highly successful International Phase of Ocean Drilling of the Deep Sea Drilling Project. The drill site, located on the south flank of the Costa Rica Rift in the eastern Equatorial Pacific, is being used to investigate a paradox within plate tectonics.

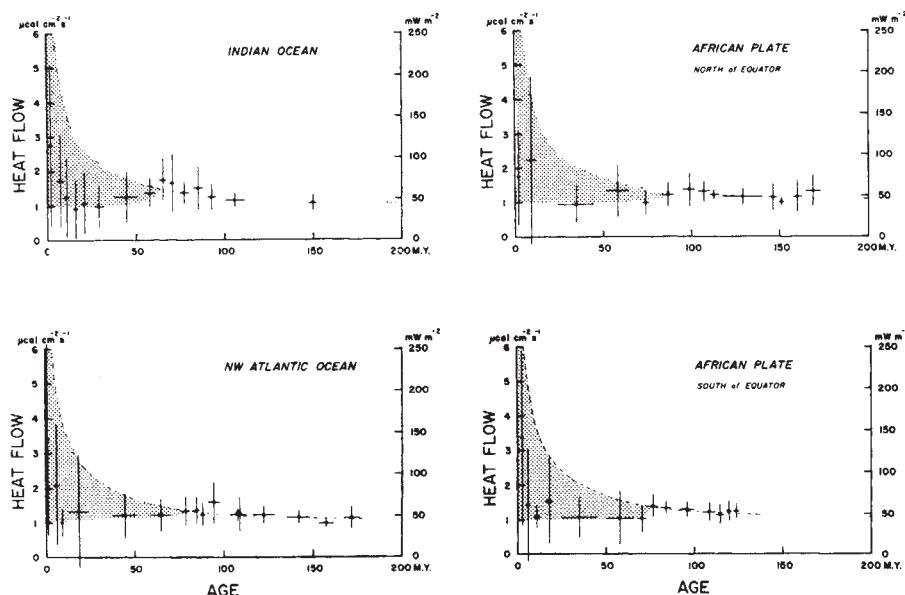
Ten years ago it was discovered that regardless of its location or tectonic setting, the depth of any normal sea floor is only a function of its age^{1,2}. This elevation change is due solely to thermal contraction accompanying cooling as crust spreads away from mid-ocean ridges. Theoretically, the heat flow measured along the top surface of the lithosphere should decrease exponentially from a maximum at spreading centres to low values in old ocean basins. Instead, as is well illustrated in a compilation prepared by M. Hobart (Lamont-Doherty Geological Observatory), observations (see the figure) show an entirely different form of cooling is taking place. Conductive heat flow is actually lowest near spreading centres, and increases to that predicted by plate tectonics as the lithosphere ages³. The 'missing heat' near midocean ridges is carried away by the hydrothermal convection of seawater through the oceanic crust and sediments⁴. With ageing, however, each plate somehow becomes sealed from this convective hydrothermal circulation and from then on the measured conductive heat flow matches that predicted by conventional lithospheric plate models.

This is no minor event since more than one-third of the entire sea floor of the

world's oceans contains a currently active convection system⁴. Every drop of the vast volume of ocean water passes through the oceanic crust in a few million years⁵. The crustal basalt consequently buffers the composition of the ocean water and is a more important source of marine chemicals than the input from rivers⁶. In addition, metallogenesis accompanying the discharge of these convective fluids from the crust into the ocean probably forms one of the Earth's major ore generation mechanisms⁷. The ocean water captured chemically within the oceanic

crust even reappears to affect man's day to day life. Volcanicity above lithospheric subduction zones is directly caused by the catalytic effect of seawater released by chemical reactions deep within the mantle⁸. Water vapour accompanying the eruption of Mt St Helens originally circulated through and was caught up in oceanic crust in the Pacific Ocean as part of a ridge flank convection cell before its eventual subduction beneath Washington and Oregon.

In 1979, during drilling legs 68–70, the *Glomar Challenger* investigated the Costa



Solid dots are regional averages of observed heat flow over the age intervals spanned by the horizontal bars. Vertical bars are ± 1 standard deviation of the data from the mean. Indian Ocean is for all plates in that ocean³. Northwest Atlantic is American Plate only. African Plate north of equator is in the Northeast Atlantic and African Plate South of the equator is in the Southeast Atlantic Ocean. Exponential decay of heat flow away from midocean ridges (dashed curve) is predicted by theoretical lithospheric plate models^{1,2}. Stipled region is the 'missing convective heat flow' component not measured by current observational techniques which only detect conductive heat transfer. With ageing, this convective hydrothermal circulation becomes sealed and we see a transition from predominantly convective to conductive heat flow. From then on, the observed heat flow matches the predicted values.

Roger N. Anderson is senior research associate at Lamont-Doherty Geological Observatory and adjunct associate professor in the Department of Geological Sciences of Columbia University, New York.

Rica Rift, the youngest known transition zone from convective to conductive sea floor heat flow. Two possible mechanisms could explain the transition — either a thick sedimentary blanket forms a hydraulic lid sealing convection within the oceanic crust and/or the basaltic 'plumbing system' becomes plugged by metamorphic alteration⁹. Only drilling would allow direct observations to be made.

Papers presented at the American Geophysical Union's spring meeting in Baltimore reported some astonishing discoveries from the integrated set of *in situ* geophysical and geochemical experiments carried out during these legs. M. Langseth (Lamont-Doherty Geological Observatory) *et al.*¹⁰ began by presenting temperature logging results which showed that seawater was being naturally drawn down the drillstring into the oceanic crust beneath 300 m of mud and chert at site 504B. The downflow rate of 6 m³ h⁻¹ was unchanged 54 days later when leg 70 reoccupied the site and it is anticipated that it will be continuing when the *Challenger* returns in November 1981.

Pore pressures 10 bars less than hydrostatic pressure as measured *in situ* across a hydraulically packed section of the wellbore were responsible for the downflow (M. Zoback, US Geological Survey and R. Anderson, ref. 11). Most of the downflow was disappearing into an aquifer 50 m below the sediment-basement interface. Direct measurements showed the basalt to be several orders-of-magnitude more permeable than the overlying sediment and chert layers¹¹. Convection in an oceanic crust with permeability of 2–40 millidarcies (determined by the measurements) can indeed account for the observed underpressures, as shown by D. Gartling (Sandia Labs) from Finite Element Convection Modelling. It thus appears that the *Challenger* penetrated through a hydraulic lid into an active convection system. Interestingly, most of the real extent of such a cell is predicted to be underpressured. Thus, many, many *Challenger* holes drilled previously in the world's oceans may have also penetrated

underpressure zones, though few measurements were ever made which might have detected such a phenomenon.

The oceanic crust does show major changes with depth, however, which must affect the form of convection. R. Von Herzen (Woods Hole Oceanographic Institution) *et al.*¹² measured a steady porosity decrease with depth from a long spacing electrical resistivity experiment. Anderson and Zoback¹³, using ultrasonic borehole televier imagery for the first time down a *Challenger* hole, found that a marked increase in seismic velocity downwards in the upper 200 m of the oceanic crust (layer 2A) was caused by replacement of seawater with clay and alteration products in fractures and voids at site 504B. R. Stephen (Woods Hole and see ref. 14) conducted an 'oblique' seismic experiment, shooting explosives from a surface ship to receivers latched into the wellbore. His results showed unusually high seismic velocities 1–2 km into the oceanic crust at this site. These may result from metamorphic alteration of the oceanic crust accompanying cellular convection.

J. Honnorez (University of Miami and see ref. 15) and M. Mottl (Woods Hole and see ref. 16) presented geochemical evidence for major chemical and alteration gradients both downhole and between four holes within one kilometre of each other at site 504B. These horizontal gradients probably result from heterogeneity caused by active hydrothermal convection. For example, Mottl predicts a major source of Ca-rich pore fluids somewhere just to the west of the drill sites. That source is probably the upwelling limb of a convection cell, the waters of which have travelled deep into the oceanic crust, reacted with basalt and then returned towards the surface of the basalt. Similarly, Honnorez reported marked increases in the 'grade' of alteration downhole with some of the 'hottest' meta-

morphic grades ever found in the drilling project (+ alc) encountered at the bottom of 504B.

It is unfortunate that 500 holes were drilled in the oceans before downhole experiments such as those using packers, televiers, borehole seismometers and long period resistivity logs were attempted, but the spectacular success of such *in situ* observations, coupled with the powerful geochemical sampling and analysis techniques now available, promise to add a truly new dimension to deep-sea drilling as a scientific tool. The return to Costa Rica will very probably produce two kilometre penetration into the oceanic crust, with a full set of *in situ* experiments.

By then, more than 10⁵ m³ of seawater will have been drawn into the oceanic basalt below the sea floor, never again to reappear in the ocean as long as the chert and sedimentary lid remain intact.

Such huge natural reservoirs have great potential as sites for the disposal of toxic wastes and offer several advantages over present proposals for land-based sites. The natural underpressures will draw seawater or wastes into the oceanic crust and if the seal is broken accidentally (by an unexpected earthquake, for example) oceanic bottom water will flow downwards into the crust rather than allowing toxic waste to escape into the ocean. Any natural heating which might accompany toxic waste burial should enhance the seal since it would convert more and more sediment into chert above the disposal cell and increase the seal's thickness. Once toxic waste was drawn into the immense oceanic crustal convection cell, its residence time would be far longer than the time for which it might endanger the environment.

Verification of these advantages is still awaited, of course, but the return of the *Challenger* to the Costa Rica Rift site should greatly help our evaluation of such deep-sea disposal sites. □

Solar physics at Oxford

from I. W. Roxburgh and C. Jordan

THE growing importance of solar physics within astronomy was illustrated by a recent meeting* at Oxford which took as its theme 'Solar Activity'. Topics ranged from the underlying causes of solar activity in the interaction between the sub-photospheric convection zone and magnetic fields to the solar flares, the highest energy manifestation of the active Sun.

Weiss (University of Cambridge) described recent theoretical work on turbulent magneto-convection. Attention

is now turning from the systematic large-scale fields, rooted deep in the convective zone, to the small-scale feature such as ephemeral active regions. Recent results suggest that these small flux loops are examples of the intermittent turbulent magnetic fields in the upper part of the convective zone, where magnetic flux has been confined to isolated ropes winding between convective eddies.

Bruzek and Schröter (Kippenheuer Institut, Freiburg) stressed the current view of solar active regions developing through the emergence and expansion of new magnetic flux. In the higher layers of the transition region and corona active region

1. Slater, J.G. and Francheteau, J. *Geophys. J. RAS* 20, 509 (1970).
2. Slater, J.G., Anderson, R.N. and Bell, M.L. *J. geophys. Res.* 76, 7888 (1971).
3. Anderson, R.N., Langseth, M.G. and Slater, J.G. *J. geophys. Res.* 82, 3391 (1972).
4. Anderson, R.N., Hobart, M.A. and Langseth, M.G. *Science* 204, 828 (1979).
5. Wolery, T.J. and Sleep, N.H. *J. Geol.* 84, 239 (1976).
6. Edmond, J. *Earth planet. Sci. Lett.* 46, 1 (1979).
7. Corliss, J.B. *et al. Science* 203, 1073 (1979).
8. Anderson, R.N., Uyeda, S. and Miyashiro, A. *Geophys. J. Res.* 44, 333 (1976).
9. Anderson, R.N. and Skilbeck, J.N. *The Sea* Vol. 7, (Interscience, 1981).
10. Langseth, M.G., Becker, K., Hobart, M. and Von Herzen, R. *EOS* 62, 308 (1981).
11. Anderson, R.N. and Zoback, M.D. *EOS* 62, 308 (1981).
12. Von Herzen, R.P., Francis, T.J.G. and Becker, K. *EOS* 62, 308 (1981).
13. Anderson, R.N. and Zoback, M.D. *EOS* 62, 308 (1981).
14. Stephen, R.A. *EOS* 62, 309 (1981).
15. Honnorez, J.J. *EOS* 62, 309 (1981).
16. Mottl, M.S. and Lawrence, J.R. *EOS* 62, 310 (1981).

*The Solar Physics Section of the European Physical Society held its Triennial meeting on 13–15 April, 1981. The meeting was preceded by a two-day workshop on 'Near Future and Plans for Solar Research'.

fields show as distinct loop structures. Substantial progress has been made in determining the temperature and density structure of these loops, and in understanding their stability (Monsignor-Fossi, Arcetri Observatory). Controversy remains concerning how much information on the heating processes can be obtained from analyses of emission line fluxes. This is an area where further high-resolution observations are required, particularly time-dependent measurements of line profiles. Maltby (University of Oslo) stressed the same point with respect to the energy balance in sunspots. Gurman (Goddard Space Flight Center) reported observing, with the UVSP instrument on the Solar Maximum Mission satellite, oscillations with periods of 120 to 180 seconds in fluxes and velocities measured in the chromosphere-corona transition region over a sunspot. It has proved difficult to detect oscillations in the quiet sun with previous satellites. Thomas and Scheuer (University of Rochester) presented results of calculations which show that oscillations could be caused by resonant behaviour due to the strong vertical trapping of wave energy. Both Maltby and Stenflo (University of Zurich) drew attention to the importance of current and future measurements of the solar irradiance from spacecraft in understanding the sunspot flux deficit.

Solar flares represent the most extreme form of activity and several reviews were given of recent results from the Solar Maximum Mission satellite. This satellite has for the first time allowed the simultaneous study of active regions and flares over a broad spectral range and thus over a range of heights and temperatures in the atmosphere. Whilst spectroscopic methods are revealing the temperature and density conditions during flares, the imaging instruments are being used to show where and when flares occur in relation to the longer-term development of particular active regions. In particular, the identification by the hard X-ray imaging spectrometer team of the source of hard X rays with the foot-points of loops, rather than with the region producing softer X-ray emission, will limit the choice of suitable models for the hard X-ray emission. Measurements of Fe K α emission caused by fluorescence and the first resolved spectra of the Fe xxvi resonance plus di-electronic satellite lines were also reported by the X-ray polychrometer team.

Before the meeting, delegates from all parts of Europe discussed their plans for the rest of the 1980s. A remarkable consensus emerged concerning important priorities including, for example, measurements of velocity fields and magnetic fields related to the development of emerging

magnetic flux; and studies of small magnetic elements in the convective super-granulation structure.

Of the ground-based programmes the most ambitious concerns the development of a large European Solar Telescope which would offer European solar physicists a high-resolution telescope situated in the Canary Islands. This project can be seen as a logical extension of the previous site-testing activities of the Joint Organization for a Solar Observatory.

Several countries will continue to participate in the USSR's Interkosmos and Prognos series of rockets and satellites, with particular interest in X-ray imaging and spectroscopy. The community within the European Space Agency is awaiting the outcome of representations to NASA concerning the US part of the two-spacecraft International Solar Polar Mission. The US spacecraft, which is due to carry out experiments to study the corona near the Sun, is

threatened by recent budget cuts.

The strength of theoretical solar physics in Europe and its close involvement in analysis of new data are particularly encouraging since they should allow a good return from participation in a variety of guest investigator programmes as opportunities arise.

There is keen European interest in two potential solar payloads for the Shuttle Spacelab: in the Solar Optical Telescope, at present under study within NASA, and in the complementary Grazing Incidence Solar Telescope, currently a potential ESA project. The future of these missions depends on the success of the Shuttle; the participants abandoned the dining hall at 7.10 on Tuesday, 14th April, to crowd round the television set and witness the safe return of the first Shuttle flight. With some relief the participants returned to continue their dinner and polish up their proposals for future experiments. □

Glueballs, anyone?

from T.F. Walsh

THE QUARK MODEL, invented in 1963 by George Zweig and Murray Gell-Mann, has gradually grown into a theory of elementary particles. It is now called quantum chromodynamics. An essential element of the theory was provided by Yoichiro Nambu in 1966. He suggested that the quarks are bound together in particles by the exchange of spin one vector particles, which have a new quantum number unlike the conventional isotopic spin, strangeness and so forth typical of the quarks. Nambu's exact model has been discarded by most physicists, but the idea remains.

The vector particles which bind quarks are now called gluons and their quantum number is called colour. Gluons interact strongly with quarks over distances of 10^{-13} cm, binding them into particles. Quarks and gluons are thought not to exist as free particles, but only inside the known 'colourless' particles, although this has yet to be proved. As well as their interaction with quarks, gluons interact strongly with each other. They should therefore bind into particles even without quarks. These particles, made entirely of gluons, have not yet been seen, but there is a strong conviction among particle physicists that they do exist. They are called bound glue states, gluonia or glueballs. They are a new type of elementary particle and it is very important to look for them.

A recent issue of *Physical Review Letters* contained two papers on glueballs, by K.

Ishikawa at DESY in Hamburg and by M. Chanowitz at the Lawrence Laboratory in Berkeley^{1,2}. Chanowitz' paper is entitled "Have We Seen Our First Glueball?" Indeed, have we? And why should the first one be seen now, after hundreds of quark states have been tabulated in the celebrated Particle Data Group tables?

If glueballs do exist, it is important to know where to look for them. According to quantum chromodynamics, heavy particles like the J/ψ and the Υ mesons can only decay first to gluons, or a photon and gluons, which then materialize into conventional mesons. The decay of the J/ψ meson into a photon and two gluons ought to be a good place to look for glueballs. If the two gluons materialize into ordinary mesons through the intermediary of a glueball, experimenters ought to see a nearly monoenergetic photon recoiling against a resonance which decays to mesons. The chain would look like this:

$J/\psi \rightarrow \text{photon} + \text{two gluons}$
 $\downarrow$
 glueball resonance
 $\downarrow$
 ordinary mesons

It has even been suggested that the two gluons should in some way maximally generate glueballs in this chain and that the production of mesons made of quarks should be much smaller than the glueball production. According to this view, one per cent of J/ψ decays would contain a glueball (there might, of course, be several glueballs visible at different masses)³⁻⁵. If only gluons

I.W. Roxburgh is Head of the Department of Applied Mathematics at Queen Mary College, London and C. Jordan is lecturer in the Department of Theoretical Physics, University of Oxford.

T. F. Walsh is a visiting scientist at CERN, Geneva.

maximally generate glueballs, that would explain why they have not been seen until now; J/Ψ and the Υ , themselves only recently discovered, would be the first good sources of glueballs. It has also been pointed out that the glueballs produced in this reaction would mostly have spin two, with the production of spinless glueballs slightly suppressed⁶.

Experimenters have studied J/Ψ decays since 1974, and decays to a photon and the eta, the eta prime and the $f(1250)$ meson (the 1250 stands for the mass in MeV) have been seen. They are produced at a rather low rate and are all well known quark states (although there is a bit of doubt about the eta prime). Recently, groups working at the SPEAR e^+e^- storage ring in Stanford, California found the decays⁷

$$J/\Psi \rightarrow K\bar{K}\pi$$

$$J/\Psi \rightarrow \eta\pi\pi$$

where the $K\bar{K}\pi$ (and, less clearly, $\eta\pi\pi$) form a narrow resonance of mass $1,440 \pm 15$ MeV and width 50 ± 30 MeV. This was a great surprise, because there is a very obscure state called the $E(1420)$ in the particle tables, and it has roughly these parameters, although there is no convincing evidence that it decays to $\eta\pi\pi$. Why should this obscure state appear together with the well known eta, eta prime and $f(1250)$?

One might guess that the new state seen in J/Ψ decay is the same as the known $E(1420)$, and that this particle is one of the long sought glueballs. However, careful examination of the experimental data shows that this interpretation cannot be correct if one does not want to upset well established ideas about the production of mesons in, for example, np collisions. Chanowitz and Ishikawa suggest that there are two mesons of mass around 1,420 MeV. One of them is listed in the data tables, and it has spin one and even parity. It decays to $K\bar{K}\pi$ but not (or seldom) to $\eta\pi\pi$. The other meson is new, seen only in J/Ψ decay (and maybe in old pp annihilation experiments at low energy). Chanowitz calls this new particle $G(1440)$, and both authors claim that it is a spin-zero glueball of negative parity. It decays to $K\bar{K}\pi$ and also to $\eta\pi\pi$. Neither author claims to prove that there are two different mesons here — what they do is to try to square the unexpected appearance of the state in J/Ψ decay with prevalent theoretical ideas.

What next? The authors suggest several steps. First, the spin and parity of the resonance seen in J/Ψ decay should be checked. If it is 0^- and the spin parity assignment of the $E(1420)$ is really 1^+ , then the one seen in J/Ψ decay is definitely a new particle. Evidence that the old $E(1420)$ and the new state decay with different rates to

$\eta\pi\pi$ would support this. Also, if the state seen in J/Ψ decay is 0^- , it will be seen in the photon-photon process

$$e^+e^- \rightarrow e^+e^- + "G(1440)"$$

in the channels $K\bar{K}\pi$ and $\eta\pi\pi$. In this process, the colliding electron and positron act as sources of photons, which then create particles. That is why the electron and positron come out of the reaction, only slightly reduced in energy having each given up a photon to make the particle. This reaction is useful because the spin-one $E(1420)$ is forbidden to appear by a theorem of C.N. Yang which says that a vector particle cannot couple to the two photons in this reaction. According to theoretical ideas, a glueball will not appear at a large rate in this reaction, but the

estimate of Ishikawa indicates that it can be seen. Finally, Chanowitz emphasizes that one should look for the proposed new $G(1440)$ in gluon jets, available in the three-jet events seen at the electron-positron storage ring PETRA in Hamburg, and in decays of the Υ meson. Again, gluons should be a good source of glueballs.

If all these experimental proposals work out and there is a new particle, $G(1440)$, made as a glueball should be, then we will indeed have seen our first glueball. And more will follow — at the very least a spin-two glueball which is expected to be produced with a significant rate in J/Ψ decay. This would provide a welcome confirmation of the ideas of quantum chromodynamics and new impulses for its further development. □

New morphological transition at the Curie temperature

from Mildred S. Dresselhaus

THE increasing sophistication of the techniques of surface science now allow the quantitative study of some long-standing problems. In a recent letter, Hamilton and Jach¹ describe a new kind of structural phase transition on low-angle nickel surfaces at the Curie temperature which gives insight into an effect discovered nearly fifty years ago². The Hedvall effect is characterized by a discontinuous change in the temperature dependence of the chemical reaction rate (Hedvall Effect I) or as a discontinuous change in activation energy (Hedvall Effect II) at the Curie temperature (T_c), the temperature of the ferromagnetic-paramagnetic phase transition. The Hedvall effects are of both theoretical and practical interest as a means to vary or control chemical reactions by external means.

The work of Hamilton and Jach¹ is an extension of previous studies by Thapliyal and Blakely³ who showed a reversible morphological transition on a stepped nickel surface using LEED (low energy electron diffraction) as a structural probe and AES (Auger electron spectroscopy) as a surface compositional probe. Hamilton and Jach demonstrated by permeability measurements that this morphological transition occurred at the Curie temperature of nickel. Both studies^{1,3} focused on vicinal surfaces, defined as single crystal surfaces in the vicinity of low-index planes. Vicinal surfaces are composed of terraces of low Miller index planes, joined by ledges a few atoms in height. Direct information on terrace widths and step heights is obtained from structure in the LEED patterns, using a kinematic theory of electron scattering from stepped surfaces⁴.

Mildred S. Dresselhaus is Professor of Electrical Engineering and Director of the Center for Materials Science and Engineering at the Massachusetts Institute of Technology.

The particular vicinal surfaces that were investigated were the $Ni(111)5^\circ[1\bar{1}0]$ and $Ni(111)10^\circ[1\bar{1}0]$ surfaces, denoting surfaces at 5° and 10° from the (111) plane in the $[1\bar{1}0]$ zone, so that the resulting stepped surfaces have (111) terraces and (110) steps. For both of these stepped surfaces, a reversible structural phase transition was observed at the magnetic ordering temperature T_c of the nickel. In the paramagnetic phase, each of the samples exhibited single-atom step heights and narrow terraces, while in the ferromagnetic phase, the $Ni(111)10^\circ[1\bar{1}0]$ sample showed a two-atom step height, and the $Ni(111)5^\circ[1\bar{1}0]$ sample a four to five-atom average step height and wider terraces.

In addition to this reversible structural transformation, an associated chemical transformation was also reported at T_c ^{1,3}. To study this transformation, clean nickel surfaces were prepared by argon-ion bombardment, heat treatment and annealing, to yield a surface exhibiting less than 1/20 of a monolayer impurity, as monitored by AES. With the nickel samples in the paramagnetic state, carbon, a dominant impurity in nickel, was found to be dispersed in the bulk of the nickel samples. In contrast, with the nickel in the ferromagnetic phase, carbon from the bulk segregated to the surface, forming a carbon coverage of ~ 0.2 monolayers just below T_c but less than 0.003 monolayers of carbon just above T_c . Assuming no entropy change at T_c due to carbon segregation at the surface, the measured carbon coverage implies a change of greater than 0.2 eV per carbon atom in the heat of segregation at the Curie point.

This change in binding of a surface atom at the magnetic phase transition is also observed in various manifestations of the Hedvall effect, such as studies of the rate of Co sublimation by Sales *et al.*⁵, the rate of carbony-

1. Ishikawa, K. *Phys. Rev. Lett.* **46**, 978 (1981).

2. Chanowitz, M. *Phys. Rev. Lett.* **46**, 981 (1981).

3. Koller, K. & Walsh, T.F. *Nucl. Phys.* **B140**, 449 (1978).

4. Brodsky, S.J. *et al. Phys. Lett.* **73B**, 203 (1978).

5. Fritzsche, H. & Minkowski, P. *Nuovo Cimento* **30A**, 393 (1975).

6. Billoire, A. *et al. Phys. Lett.* **80B**, 381 (1979).

7. Scharre, D. *et al. Phys. Lett.* **97B**, 329 (1980).

lation of Ni by Mehta *et al.*⁶ and the rate of Ni oxidation by Sales *et al.*⁷. In all these studies it was found that more energy must be expended to remove a surface atom from a ferromagnetic substrate than from a paramagnetic one. The observations are explained by a lowering of the internal energy of the system as magnetic order is established at the onset of the ferromagnetic state. To overcome this additional binding, more energy must be supplied to remove a surface magnetic atom from a ferromagnetic substrate.

The sublimation of a magnetic species such as cobalt represents an especially attractive process for the study of the magneto-chemical Hedvall effect because of the simplicity of the reaction and because of the removal of the reaction product once the reaction is completed. On the other hand, technical difficulties associated with the high temperature experiment ($T_c \sim 1,400$ K for Co) had to be overcome to make quantitative measurements. The change in the apparent activation barrier for sublimation on passing through the Curie temperature for Co is 18 kcal per mole or about 0.8 eV per atom. By analysing the temperature dependence of the sublimation rate below and above T_c , Sales *et al.*⁵ were able to show that an additional binding energy, $E_m(T)$, was introduced in the ferromagnetic state and associated with the spontaneous ordering of the magnetic moments, $E_m(T) \sim [M(T)/M(0)]^2$, where $M(T)$ is the spontaneous magnetization at temperature T . Because of the rapid change in $M(T)$ near T_c , a large apparent change in activation energies is found near T_c for physically reasonable values of $E_m(0)$, such as 0.2 eV per atom for Co. The magneto-chemical effects associated with this sublimation process have been qualitatively explained in theoretical models by Suhl⁸ and by D'Agliano and Huberman⁹.

Another chemical reaction for which the reaction product leaves the surface on completion of the reaction is the nickel carbonylation reaction, whereby nickel reacts with CO gas to form gaseous nickel carbonyl. This reaction has the further attractive features of a convenient temperature range ($300 < T < 440$ K) and a Curie temperature controllable by dilution of the nickel with copper (which does not react with CO in this temperature range).

Just as for the sublimation reaction, the nickel carbonylation reaction also shows a change in the apparent activation barrier at the Curie temperature, which in this case is about 0.4 eV per atom. This experiment provides further evidence that an additional activation barrier E_m is associated with the ferromagnetic state, insofar as E_m decreases with increasing Cu concentration such that $E_m \rightarrow 0$ for the Cu concentration where $T_c \rightarrow 0$.

Extensive studies of the Hedvall effect have also been carried out on the oxidation of nickel⁷ and of iron¹⁰, yielding general agreement with the basic conclusions of the above magneto-chemical experiments. These oxidation experiments have, however, been more difficult to interpret

quantitatively because of the accumulation of the reaction product at the free oxide surface and the concomitant presence of the additional magnetic barrier at the internal metal-oxide interface.

The identification of a morphological surface transformation at the Curie temperature T_c suggests new insights into the magneto-chemical Hedvall effect because of the enhanced chemical activity at surface steps and kinks. At this point one can only speculate that the discontinuity in the reaction rate at T_c (Hedvall Effect I) might be associated with a morphological change in step size, while a discontinuity in the activation energy or in the derivative of the reaction rate (Hedvall Effect II) corresponds to no morphological change. □

Sir Isaac Newton and his madness of 1692 and 1693

from Milo Keynes

BY NO MEANS had Sir Isaac Newton a perfect character. William Whiston who succeeded him in his Lucasian professorship at Cambridge said of him that "he was of the most fearful, cautious and suspicious temper"; John Flamsteed, the first Astronomer Royal, declared he was "insidious, ambitious, and excessively covetous of praise, and impatient of contradiction". Even his firm friend, John Locke, wrote that "He is a nice" (that is, difficult and over-precise) "man to deal with, and a little too apt to raise in himself suspicions where there is no ground". There are, however, many instances of his kindness and generosity to others.

Newton had an abnormal dread of controversy, and he wrote to the experimental scientist, Robert Hooke, that "There is nothing I desire to avoid in matters of philosophy more than contention, nor any kind of contention more than one in print". Newton was neurotic, and as John Maynard Keynes wrote¹: "His deepest instincts were occult, esoteric, semantic — with profound shrinking from the world, a paralysing fear of exposing his thoughts, his beliefs, his discoveries in all nakedness to the inspection and criticism of the world. He parted with and published nothing except under the extreme pressure of friends. Until the second phase of his life (as Master of the Mint in London, and President of the Royal Society), he was a wrapt, consecrated solitary, pursuing his studies by intense introspection with a mental endurance never equalled". Besides all this, he underwent a period of severe emotional and mental disturbance

with sleeplessness and loss of appetite during 1692 and 1693.

In 1693, Newton wrote to Samuel Pepys: "I am extremely troubled at an embroilment I am in, and have neither ate nor slept well this twelve month, nor have I my former consistency of mind, but am now sensible that I must withdraw from your acquaintance, and neither see you nor the rest of my friends any more, if I may but leave them quietly". The same year he wrote to Locke: "Being of the opinion that you endeavoured to embroil me with women and by other means I was so much affected with it as that when one told me that you were sickly and would not live I answered twere better you were dead. I desire you to forgive this incharitableness". Both Pepys and Locke appear to have recognised that Newton's mind was deranged and maintained their friendship for him, and in other letters it is clear his memory was impaired.

It was with the publication of *The Correspondence of Sir Isaac Newton*² in 1961 that the period of irrationality could be more fully recognised. Newton was born in 1642 and died in London in 1727. He matriculated at Trinity College, Cambridge in 1661, became Lucasian Professor in 1669, and Fellow of the Royal Society in 1672. He left Cambridge in 1696 to become Master of the Mint in London where "with Pepys and William Lowndes (Secretary to the Treasury) he became one of the greatest and most efficient of our civil servants. He was a very successful investor of funds, surmounting the crisis of the South Sea Bubble, and died a rich man" (Keynes¹). He was elected President of the Royal Society in 1703, being annually re-elected for the next 24 years, and in 1705 he was knighted by Queen

1. Hamilton, J.C. and Jach, T. *Phys. Rev. Lett.* **46**, 745 (1981).
2. Hedvall, J.A., Hedin, R. and Persson, O. *Z. physik. Chem.* **B27**, 196, (1934).
3. Thapliyal, H.V. and Blakely, J.M. *J. Vac. Sci. Technol.* **15**, 600 (1978).
4. Henzler, M. *Surf. Sci.* **36**, 109 (1973).
5. Sales, B.C., Turner, J.E. and Maple, M.B. *Phys. Rev. Lett.* **44**, 586 (1980).
6. Mehta, R.S., Dresselhaus, M.S., Dresselhaus, G. and Zeiger, H.J. *Phys. Rev. Lett.* **43**, 970 (1979).
7. Sales, B.C. and Maple, M.B. *Phys. Rev. Lett.* **39**, 1636 (1977); Sales, B.C., Maple, M.B. and Vernon, F.L. III *Phys. Rev.* **B18**, 486 (1978).
8. Suhl, H. *Phys. Rev.* **B11**, 2011 (1975).
9. Galleani D'Agliano, E. and Huberman, B.A. (private communication).
10. Sales, B.C., Cabrera, A.L. and Maple, M.B. *Solid State Commun.* **30**, 119 (1979).

Milo Keynes is in the Department of Anatomy, University of Cambridge.

Anne — although for political, rather than scientific reasons.

In London, "he reigned as the most famous man of his age, of Europe, and — as his powers gradually waned and his affability increased — perhaps of all time, so it seemed to his contemporaries" (Keynes¹), but in the more than 30 years he lived there after leaving Cambridge, his illness of 1692 and 1693, which consisted of insomnia, loss of appetite, loss of memory, melancholia (depression), delusions of persecution, and possibly a trembling in his writing, appears to have been forgotten. After 1693, he still produced copious writings on his mathematical work, on lunar motion, atmospheric refraction and on resisted projectile motion². He made the corrected Proposition X of the second book of *Principia Mathematica* in his seventieth year, and recast his theory of the propagation of light between 1705 and 1715. Besides the comment that he spoke very little in company after 1693, there was something rather languid in his look and manner in those years, as shown in a portrait, circa 1689, by Sir Godfrey Kneller in the possession of the Earl of Portsmouth.

Various explanations have been suggested for his illness, predominantly psychological, such as the shock at the death of his mother in 1679, a fire in which he lost some papers (though this occurred far earlier than 1692), his failure to gain various posts such as the Provostship of King's College, Cambridge (which he did not get for the reason he did not attend Eton College as a boy) or the Mastership of the Charterhouse, or simple exhaustion from writing the *Principia* which was, however, finished in 1687. Clearly, none of these is at all satisfactory, but recently it has been suggested^{4,5} that Newton's nervous breakdown might be explained by mercury poisoning arising from his experiments in alchemy and in making mirrors. Metallic mercury absorption by inhalation as mercury vapour, or by ingestion, can produce neurological, psychiatric, oral and renal signs, and these are reversible when exposure ceases. Mercury is stored in the brain, kidneys and liver, and over a period of time after the poisoning has stopped will continue to appear in hair and nails.

Newton's interest in alchemy was first shown in 1667 when he purchased the six volumes of Zetzner's *Theatrum Chemicum*, as well as chemicals and apparatus, though the first dated chemical

experiment in his notebooks was in 1678, with the last in 1696 before he moved to London. He gave up his experiments for nearly two years from May 1684 till April 1686, the period in which he wrote the greatest part of his *Principia*, but during 18 years he carried out several hundred chemical experiments, with metals playing a large part. It is not always sure from the note books whether the experiments were actually performed, or were just thought of, and often it is not clear any longer what were the chemicals used. It is, however, well recorded that he would heat metals, or their ores, with salts in open vessels, breathing in the fumes and sometimes tasting the formed products. He used to sleep in his 'laboratory' by the fire, which was kept burning for weeks on end while the experiments were going on, and clearly exposed himself to the dangers of metal poisoning.

Newton's great niece, Catherine Conduit, married John Wallop, first Earl of Portsmouth, and as a result Newton's relics and manuscripts mainly passed to the Portsmouth archives. Spargo and Pounds⁵ obtained some of Newton's hairs from the Earl of Portsmouth as well as from Trinity College, Cambridge, and analysed them for mercury, gold, arsenic, antimony and lead. We do not know when the hairs were collected (though we do know Newton kept his hair short in older age which increase the chance that those analysed derived from a time nearer to the experiments), but

raised levels of the metals were found, especially of mercury and lead, and this would be most unlikely, even allowing for contamination from any vessels in which the hairs were stored, in ordinary hairs. It may be added that although some of Newton's symptoms could be caused by lead poisoning, all could be caused by mercury poisoning. We can presume that Newton showed *erethismus mercurialis*, that is, a timidity especially in the presence of strangers, but we do not know for certain whether he showed the tremor of mercury poisoning (Hatters' shakes, a condition found among those using mercury in the felt hat industry and from which Lewis Carroll culled his 'Mad Hatter') as his handwriting did not decisively deteriorate during his breakdown.

In his recent lengthy biography, Westfall⁶ suggests that Newton's symptoms were simply due to overwork in finishing the *Principia* in 1687, five years before their onset. This seems an unlikely explanation in view of the delusions Newton suffered and the raised levels of mercury and lead in his hair. The epitaph remains with Newton, who shortly before he died, said: "I do not know what I may appear to the world; but to myself I seem to have been only like a boy, playing on the seashore, and diverting myself, in now and then finding a smoother pebble or a prettier shell than ordinary, while the great ocean of truth lay all undiscovered before me". □

Are transport proteins porous?

from N. Michael Green

IN the absence of reliable methods for determining molecular size in the lipid bilayer, the state of aggregation of proteins in membranes has proved difficult to determine. A variety of indirect methods have led to conflicting interpretations. Even when the state of aggregation has been established, further experiments are required to clarify its relation, if any, to the function of the protein and even if the oligomeric structure is proved to be essential, the further interesting question of the inter- or intra-subunit location of a transport channel requires an experimental answer.

The difficulty of evaluating the evidence is emphasized by the somewhat conflicting conclusions of two recent reviews. The more wide-ranging article, (Klingenberg *Nature* 290; 449, 1981) considers thirteen membrane proteins or complexes of proteins for which there is evidence for an

oligomeric structure (these include the ADP/ATP carrier, (Na⁺ + K⁺)- and Ca²⁺-ATPases, anion carrier (band III), glucose carrier, porin, bacterial rhodopsin, acetylcholine receptor and various cytochrome complexes). The evidence, derived mainly from sedimentation experiments in non-ionic detergents and from cross-linking experiments in membranes, shows that these proteins are all either dimeric, or trimeric, although when the dust settles around the acetylcholine receptor it is likely that each unit of the dimer will be found to be a pentamer.

Klingenberg emphasises the useful conclusions that can be drawn from consideration of symmetry. The vectorial activities of transport proteins imply the absence of symmetry axes from the plane of the membrane and this eliminates dihedral structures from consideration. The popular tetrameric structure for soluble, globular proteins is therefore unlikely to be found among the proteins of membrane transport which must possess cyclic symmetry and a transmembrane

1. Keynes, J.M. *Newton, the Man* (The Royal Society Newton Tercentenary Celebrations) (Cambridge University Press, 1947).

2. *The Correspondence of Sir Isaac Newton* 3 (Cambridge University Press, 1961).

3. Whiteside, D.T. Personal communication, 1980.

4. Johnson, L.W. & Wolbarsht, M.L. *Mercury Poisoning: a Probable Cause of Isaac Newton's Physical and Mental Ills* (Notes and Records of the Royal Society of London 34; 1, 1979).

5. Spargo, P.E. & Pounds, C.A. *Newton's 'Derangement of the Intellect': New Light on an Old Problem*. (Notes and Records of the Royal Society of London 34, 11; 1979).

6. Westfall, R.S. *Never at Rest: a Biography of Isaac Newton* (Cambridge University Press, 1981).

N. Michael Green is in the National Institute for Medical Research, Mill Hill, London.

symmetry axis. Such a structure will be stable in a lipid environment only if the intersubunit links include a strong polar component, the symmetry axis providing a plausible location for a polar channel for translocation of substrates.

The accommodation of ligand binding sites and a gating mechanism in a dimeric channel raises a number of interesting questions and leads to the conclusion that there are likely to be only half as many active sites as subunits. Klingenberg develops this argument around the ADP/ATP exchange protein of mitochondria which his laboratory has clearly shown to exhibit such half-of-sites reactivity towards the mutually exclusive inhibitors bongkrekate and carboxyatractyloside. The evidence for other carriers is less clear cut and Klingenberg concludes that while the $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ and the glucose carrier of erythrocytes show half-of-sites reactivity, the anion transport protein, the β -galactoside carrier and bacterial rhodopsin probably have separate channels associated with each subunit.

Kyte (see this issue of *Nature*, p.201) directs his attention mainly to the $(\text{Na}^+ + \text{K}^+)\text{ATPase}$, though he considers his arguments are likely to apply with equal force to the functionally homologous $\text{Ca}^{2+}\text{-ATPase}$ and $(\text{H}^+ + \text{K}^+)\text{ATPase}$. He takes a sceptical attitude to the evidence for dimeric structure, pointing out the problems associated with: (1) the large corrections which have to be made for bound detergent and liquid, (2) the underestimation of the molecular weights of intrinsic membrane proteins by gel electrophoresis, and (3) the interpretation of the results of cross-linking experiments in membranes, where intermolecular distances are often very small. Nevertheless he concludes that on balance the evidence does favour a dimeric $(\text{Na}^+ + \text{K}^+)\text{-ATPase}$.

The evidence for half-of-sites reactivity is much less convincing, particularly in the light of two recent papers in which the equivalent weight of the ATPase towards ouabain and towards TNPATP has been critically re-evaluated as $180,000 \pm 20,000$. This figure, consistent with one site per 2β protomer, results mainly from use of a more objective method for determining the protein concentration (amino acid analysis) in place of the uncalibrated Lowry method. He also points out that the anti-cooperative kinetics of all three transport ATPases with respect to ATP does not necessarily imply interaction between the two sites on a dimer; it is equally consistent with a monomer in which the ATP site can exist in interconvertible low- and high-affinity conformations. In view of the elusive behaviour of apparent half-of-sites reactivity of globular proteins under critical examination, stronger evidence is required to substantiate it for the $(\text{Na}^+ + \text{K}^+)\text{ATPase}$.

Doubts also surround the $\text{Ca}^{2+}\text{-ATPase}$ of sarcoplasmic reticulum. Although it possesses one ATP site per peptide chain only half of them appear to be phosphorylated in most membrane preparations. Moreover, recent evidence (Pick and Karlisch *Biochim. biophys. Acta* 626, 255) suggests that one mole of covalently bound fluorescein isothiocyanate can block the activity of two peptide chains. On the other hand, the monomer in non-ionic detergent is a fully active ATPase showing steady-state kinetics which differ little from those

of oligomeric preparations.

Of the well characterised transport proteins showing apparent half-of-sites reactivity, only the ADP/ATP exchange protein survives unchallenged; even here the intersubunit channel is no more than the most plausible hypothesis. There is firm structural evidence for the location of the channel in bacterial rhodopsin which functions as a monomer and contains a plausible intrasubunit channel. For the remainder the definitive experiments have still to be devised.

Acid precipitation — a new study from Norway

from J.N.B. Bell

IN THE 1960s the destruction of freshwater fish stocks and the impairment of forest productivity in Scandinavia were blamed on an increase in the acidity of water and soils. It was believed that the increase resulted from 'acid rain' — precipitation containing sulphuric acid derived from the oxidation of SO_2 emitted by the more industrialized European nations. Subsequently there have been similar claims elsewhere, particularly in the north-eastern United States and Canada.

Public concern has nowhere been more evident than in Norway, where in the mountainous regions of the south, an accelerating decline in fish stocks has led to the extinction of fish populations in an area of 13,000 km², with severe problems over another 20,000 km². This stimulated the establishment of a major Norwegian interdisciplinary research programme into the biological effects of acid precipitation, which has recently terminated after eight years. The numerous reports published on the findings this project have now been summarized (Overrein, Seip and Tollan *Acid precipitation — effects on forest and fish*. Final report of the SNSF project 1972–1980).

It is clear from the final report that the SNSF project has made a major contribution towards the understanding of the relationships between the chemistry of precipitation, soils and aquatic ecosystems. The main problems in determining the ecological significance of acid precipitation lie in extrapolating the results of short investigations to the long-term effects of relatively small changes in soil and water chemistry. It has now been established beyond doubt that the precipitation in southern Scandinavia has become more acidic as a result of long-distance transport of air pollution, with

nitrogen oxides contributing an increasing proportion of the acidity. The inputs of H^+ , SO_4^{2-} and NO_3^- are readily estimated by routine analysis of precipitation samples collected at many sites in Norway. However, some uncertainty remains as to the importance of the dry deposition of gaseous and particulate sulphur and nitrogen compounds directly onto vegetation and soils, which may make a substantial contribution to total deposition in some areas.

The report admits that there are few reliable records of the acidity of lake water before 1950 but the marked shift since then towards more acid conditions in areas subject to acid precipitation is well documented. Only five to ten per cent of precipitation in southern Norway falls directly onto lakes and rivers. Thus a thorough understanding of the chemical changes occurring in water when it comes into contact with solids and vegetation is of prime importance when considering the ultimate effect on aquatic ecosystems. A key component of the SNSF project has been the assessment of annual precipitation input and ground-water output into waterways of major ions in selected catchments. It is now apparent that, as well as increased acidity, the mobilization of aluminium from soils by acid water is a major factor in the destruction of fish stocks. The uptake of salt via the gills, necessary to maintain the osmotic balance of the fish, is impaired by acid conditions and is further aggravated by aluminium stimulating the secretion of mucus.

Particularly severe problems are experienced in southern Norway, because a large proportion of the precipitation falls as snow. Pollutants accumulate within the snow-pack during the winter and are subsequently released very rapidly during the short period of snow-melt in the spring, resulting in a flush of acid water into lakes. Knowledge of the hydrology of snow-melt is limited but the SNSF project has

J.N.B. Bell is at the Centre for Environmental Technology, Imperial College of Science and Technology, London.

examined the chemical content of successive fractions of melt-water: up to eighty per cent of the pollutant content of a snow-pack may be released in the first thirty per cent of the melt-water, so increasing the threat to the sensitive larval stages of salmon and trout.

The effect of acid precipitation on terrestrial vegetation is less certain. Despite an intensive research programme, the SNSF project provides little support for earlier predictions of massive losses in timber production through leaching of nutrients from soils and direct effects on trees. Most experiments demonstrating adverse effects on soils and plant growth used artificial rain with unrealistically low pH levels. Indeed, in several cases stimulation of plant growth has been observed. The situation is complicated because acid precipitation can increase the availability of plant nutrients by providing inputs of sulphur and nitrogen from the atmosphere and releasing basic cations by weathering of minerals in the soil. It is rather unfortunate that this generally well balanced report implies that acid precipitation must be assumed to reduce tree growth, in the absence of proof to the

contrary.

The causal relationship between acid precipitation and the decline of fish stocks has been questioned by Rosenqvist (*Sci. Total Environ.* 10, 39; 1978), who drew attention to the impact of changes in catchment land use on the chemistry of aquatic systems. During the last one hundred years, many of the Norwegian upland farms, used for summer grazing, have been abandoned and in some cases replaced by conifer woods, which would be expected to accelerate natural soil acidification processes. The SNSF programme has found no consistent pattern which could explain lake acidification on the basis of land use changes. However, an investigation in central Scotland has shown a clear effect of afforestation in increasing stream acidity (Harrison and Morrison in *Ecological Impact of Acid Precipitation*, eds Drabl and Tollan, SNSF Project, 312; 1980). The report concedes that afforestation may enhance the acidification caused by precipitation. This remains an area of uncertainty and there is a clear need for a better understanding of the relative importance of different sources of

acidification at sites where fish stocks are under stress.

The eight-year SNSF project has clarified many aspects of the acid precipitation problem. The claims that reduction of fish stocks has been caused by acid precipitation have generally been substantiated, although questions remain concerning the contribution from changes in land use. In contrast, the long-term effect of acid precipitation on forest productivity is far from understood and intensive studies have failed to reveal any clear trends. The SNSF project has predicted that a reduction of at least seventy per cent in H^+ and SO_4^{2-} concentrations in precipitation is necessary to restore all lakes to their previous condition: this would involve international changes in energy policy which would encounter severe political obstacles. So far as action within Scandinavia is concerned, attempts to improve the water quality by the addition of lime, followed by restocking with fish, have achieved only limited success and a profitable course may be an extension of a preliminary SNSF programme for breeding strains of trout and salmon tolerant to acid conditions. □

100 years ago

"Anthropology: an Introduction to the Study of Man and Civilisation, by Edward B. Tylor, D.C.L., F.R.S. With Illustrations. (London: Macmillan and Co., 1881.)"

To those readers whose knowledge of ethnology or anthropology has been derived from a perusal of Prichard's "Natural History of Man," or the compilations of Wood, Brown, Peschel, or Brace, the present work will present a surprising amount of freshness and originality. They will in fact find themselves introduced to a new and very captivating science.



Andaman Islanders

The first chapter contains a brief sketch of what we learn from history, archaeology, and geology, as to man's antiquity and early condition; and in the next we are shown man's relation to the lower animals both in bodily structure and mental characteristics. The numerous remains now discovered of prehistoric man, and of his works, dating back to an undoubtedly vast antiquity, show us in no case any important deviation from the existing human type, nor any indication that his mental status was lower than (if so low as) that of many living races. At the same time the increasing rudeness of his implements as we go back, undoubtedly indicates that we have

made some approach towards the period when he first emerged from the purely brute state and became "a tool-using animal."

In the next chapter we have an excellent sketch of the chief races of man copiously illustrated by portraits, mostly from photographs and very characteristic. Among the best are those of the Andaman Islanders and the Dyaks, which we here reproduce.

The four chapters on Language, are exceedingly interesting and instructive, especially the account of the gesture language and the illustrations of how connected stories may be told to the deaf-and-dumb quite independently of any knowledge of alphabetical or even verbal signs. In treating of the origin of language Mr. Tylor doubts the sufficiency of the theory that emotional, imitative, and suggestive sounds were the basis on which all languages were founded, though he gives tolerably full illustrations of how roots thus obtained became modified in an infinite variety of ways to serve the growing needs of mankind in expressing their wants or their feelings. Putting aside all mere representations of animal sounds — as the *whinny* of the colt, the *mew* of the cat, or the *bleat* of the sheep — consider how clearly do such words as *slide*, *glide*, and *wave* imply slow and continuous motion, the movement of the lips while pronouncing the latter word being a perfect double undulation. How curiously do the tongue and palate seem to be pulled apart from each other while pronouncing the words *glue* or *sticky*. How marked is the contrast between the harsh consonants used to express *rough*, *rugged*, and *gritty*, as compared with the soft flow of sounds in *smooth*, *oily*, *even*, *polished*. Among the Malay races, for instance, in words for *large* we find a prevalence of broad sounds involving a wide opening of the mouth, as *busar*, *baké*, *bagut*, *lamu*, *elamo*, *ilahé*, *erámei*, *aiyuk*, *máina* — and for *small*, words that are pronounced quickly and with slight opening of the lips, as *kichil*, *chili*, *kidi*, *kôï*, *roit*, *kemi*, *anan*, *kiiti*, *fek*, *didiki*, all taken from languages of the Malay Archipelago.

The five following chapters treat of the Arts of Life, a subject which Mr. Tylor has to a great extent made his own, and which he discusses in a very interesting manner. The



Dyaks

doctrine of development in the arts is however somewhat strained when it is implied that the modern gun is an outgrowth of the South American or Indian blow-tube; while the origin of bank notes, and the account of the rise and progress of mathematics are hardly anthropology.

The next two chapters discuss the ideas of savage man as to the spirit-world, and the origin and development of myths; while the final chapter gives an admirable sketch of man as a social being, and of the development of that complex organism, Society. This thoughtful chapter cannot be epitomised, but the reader will find in it much curious information as to the sources of many of the customs, laws, and observances of civilised life, which are shown to be often traceable among the lowest savages.

ALFRED R. WALLACE

From *Nature* 24, 14 July, 242, 1881.

REVIEW ARTICLE

Molecular considerations relevant to the mechanism of active transport

Jack Kyte

Department of Chemistry, D-006, University of California, San Diego, La Jolla, California 92093, USA

A small group of closely related proteins is responsible for all active transport in animal cells, and inorganic cations are the only substances transported by these enzymes. They share a common kinetic mechanism in which two fundamental conformations participate, each receiving and dispatching substrates from its unique side of the membrane. During transport, the cations must pass through their enzyme to cross the membrane and intense interest is currently focused on the possibility that the path which they follow lies within the interface between two discrete subunits in a dimeric structure. Although 'half-of-sites' behaviour, consistent with this hypothesis, has been reported, it is now known that systematic errors were responsible for this mistaken conclusion. The number of protomers which comprise a functional unit of active transport has not been determined.

THE plasma membrane of an animal cell forms the permeability boundary which separates the cytoplasm from the external environment. A diverse collection of metabolites is transported across the plasma membrane by a group of transport proteins, each of which mediates the flux of a specific substrate or group of substrates. Enzymes responsible for the active transport of sodium and potassium¹, calcium^{2,3}, and protons and potassium⁴ have been purified to homogeneity. The minimum stoichiometries for the chemical equilibria catalysed by these enzymes are, respectively:



where i and o indicate the cytoplasm and extracytoplasmic spaces, respectively⁴⁻⁶. In broken membrane fragments or preparations dispersed in detergents, these reactions are expressed as a sodium- and potassium-dependent ATPase ((Na⁺ + K⁺)ATPase), a calcium-dependent ATPase (Ca²⁺-ATPase) and a proton- and potassium-dependent ATPase ((H⁺ + K⁺)ATPase). The movements of the cations in each direction and the hydrolysis of the MgATP are coupled obligatorily so that the free energy from the hydrolysis of ATP is converted into electrochemical gradients of the cations. The similarity of these three chemical reactions is quite striking.

No other enzymes which reside in the extramitochondrial membranes of animal cells and which can convert metabolic energy directly into the work required to move molecules against a concentration gradient have been described. In addition, every other known example of net extrusion or accumulation of any substance against its concentration gradient across an animal cell plasma membrane involves the coupling of that movement to the electrochemical gradient of sodium created by (Na⁺ + K⁺)ATPase. In other words, the three chemical equilibria shown above describe all known examples of active transport in animal cells. Therefore, when active transport is used as a term here, it will refer only to the reactions catalysed by (Na⁺ + K⁺)ATPase, Ca²⁺-ATPase and (H⁺ + K⁺)ATPase.

These enzymes, as well as a membrane-bound ATPase in fungi responsible for active proton extrusion⁷, are very closely related to each other. They each contain a large polypeptide chain whose apparent length, from SDS-polyacrylamide gel analysis, lies between 900 and 1,200 residues^{1-4,8}. In each case

this polypeptide is the phosphorylated intermediate associated with catalysis^{3,9-12}. Additional similarities have been noted between Ca²⁺ ATPase of the sarcoplasmic reticulum and (Na⁺ + K⁺)ATPase of the plasma membrane. Both proteins have an identical sequence around the aspartic acid residue which is phosphorylated during turnover¹³, and both have membrane-affiliated regions distributed throughout the catalytic subunit¹⁴⁻¹⁶. Finally, the phosphorylated aspartic acid in both proteins is located in the same region of the catalytic subunit, 200-400 residues from the amino terminus^{16,17}. Although the other enzymes which catalyse active transport, (H⁺ + K⁺)ATPase and the fungal H⁺-ATPase, have yet to be studied as carefully, it is reasonable to assume that they will share these similarities. These correlations suggest that all the enzymes which catalyse active transport share a common ancestor. Thus, conclusions drawn from any one of these enzymes probably apply to all of them. In particular, they all must have very similar kinetic mechanisms which proceed through analogous intermediate steps and all must have equivalent tertiary and quaternary structures, positions within the bilayer and cation channels.

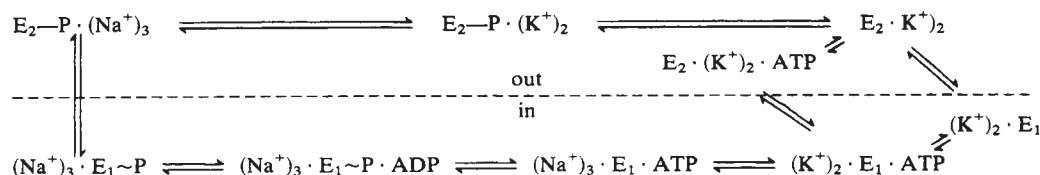
It is unclear how these enzymes move cations across the membrane or how they couple the movement of the cations to ATP hydrolysis. The answers to these questions are likely to involve molecular explanations of at least two conformations which these enzymes can assume on the one hand, and on the other, of the minimum unit necessary for enzymatic function.

Kinetic mechanism

Various aspects of the sequence of kinetic steps through which (Na⁺ + K⁺)ATPase passes during its turnover have been examined both in membrane fragments and in the ATP-dependent, ouabain-inhibited, active cation fluxes of sealed systems. (Na⁺ + K⁺)ATPase can be covalently phosphorylated on an aspartate residue by the γ phosphate of MgATP¹⁸; the phosphorylated intermediate is formed in the presence of sodium but hydrolysed in the presence of potassium¹⁹. This acyl phosphate has rates of formation and decay which are rapid enough for it to be an intermediate in the kinetic sequence of the reaction of the enzyme. When the rate of its forward decomposition is constant, the rate of overall turnover is directly proportional to its steady-state concentration²⁰. These experiments clearly establish that the enzyme is phosphorylated and dephosphorylated each time it turns over, and they divide active transport into two distinct steps.

Although the distinction between phosphorylated and unphosphorylated enzyme is a simple one involving the presence or absence of a covalent bond, there also seem to be at least two unique conformations of the native enzyme, the definitions of which are rather more ambiguous. Nevertheless, it is believed that the conformational changes which interconvert these states are more fundamental than phosphorylation and are responsible for the actual movement of the respective cations from one side of the membrane to the other. The two conformations, designated E_1 and E_2 , differ in their affinity for ATP^{21,22}, their intrinsic fluorescence spectra²³, susceptibility to trypsin^{17,24,25} and the response of their phosphorylated intermediates to ADP and K^+ (ref. 26). Transitions between E_1 and E_2 are detected by changes in any one of these properties. Both *N*-ethylmaleimide alkylation²⁷ and oligomycin²⁸ prevent these transitions.

Although there is still extensive disagreement about the details of the process, the following scheme is at least consistent with the experimental observations^{22,23,29,30}.



Some of the steps in this scheme are proposed as the result of kinetic inference, but most have been directly observed and rate constants or equilibrium constants have been assigned to them. In normal circumstances the cycle operates in the clockwise direction, extruding sodium and accumulating potassium. Kinetic experiments indicate that the transition from $E_1 \sim P$ to $E_2 - P$ requires sodium to be bound to the enzyme, whereas the transition between E_2 and E_1 requires potassium. This suggests that E_1 is the inward-facing form of the enzyme which normally releases potassium as a product and receives sodium as a substrate, and E_2 the outward-facing form which releases sodium as a product and receives potassium as a substrate. Therefore, it is during the transition between these two forms that the cations traverse the plasma membrane. Kinetic experiments with Ca^{2+} -ATPase^{31,32} and $(H^+ + K^+)$ -ATPase³³ have been interpreted in terms of a mechanism completely included within the more detailed mechanism which is currently available for $(Na^+ + K^+)$ -ATPase. No counter ion, however, has been identified which would have the same role in Ca^{2+} -ATPase that K^+ does in the mechanism of $(Na^+ + K^+)$ -ATPase and $(H^+ + K^+)$ -ATPase. It should be emphasized that this mechanism is able to explain^{22,34} the negative cooperativity displayed in the MgATP kinetics of Ca^{2+} -ATPase³⁴, $(H^+ + K^+)$ -ATPase³³ and $(Na^+ + K^+)$ -ATPase³⁵ without any need to postulate half-of-sites behaviour or inter-subunit interactions.

Molecular structure

While these kinetic observations were being accumulated, another series of experiments were being done to provide evidence for a particular structural explanation of the transition between E_1 and E_2 . These experiments evolved from the following theoretical argument³⁶. In a lipid bilayer all protein molecules of the same sequence are inserted so that they point in the same direction^{37,38}. As the same area on the surface of each molecule, which is defined by a preponderance of aliphatic amino acid side chains³⁹, interacts with the boundary lipid⁴⁰, all folded polypeptide chains of the same sequence float at the same depth. With few exceptions, when two or more identical polypeptide chains combine to form a multi-subunit complex, they form a closed structure by arranging themselves about a rotational axis of symmetry⁴¹. In the membrane, this axis of symmetry is necessarily normal to the plane of the bilayer. The simplest arrangement of subunits in a membrane-bound

multimer, which has evolved under all these constraints, is a dimer whose dyad axis runs through the membrane. Furthermore, the only significant channels which pass through the centre of protein molecules lie between subunits and are often found centred on axes of symmetry. The true dyad of haemoglobin is the earliest and most striking example of this⁴². These consequences, which result when the rules that govern the symmetry of oligomeric proteins⁴¹ are applied to the two-dimensional environment of the phospholipid bilayer^{36,43}, have recently been reviewed by Klingenberg⁴⁴.

It is known that active transport of sodium and potassium through membranes does not involve the rotational or translational diffusion of $(Na^+ + K^+)$ -ATPase across the membrane⁴⁵. Thus it has been proposed that the enzyme spans the plasma membrane³⁶ and forms a channel across it through which the cations pass. In this scheme the pathway for the cations lies through the centre of the enzyme molecule, and is formed by the juxtaposition of two subunits of the enzyme and located along the dyad axis between them. The enzyme alternates between

two conformational states, E_1 and E_2 , during turnover. These conformations differ from each other primarily in small structural alterations in the geometry of the channel; E_1 provides access to the cation compartment from the inside of the cell, E_2 from the outside. During the conformational changes which lead to net transport, each cation remains in the compartment and then departs into the medium on the side of the plasma membrane opposite to that from which it originally came. This hypothesis for the mechanism of $(Na^+ + K^+)$ -ATPase³⁶, and by analogy for the other enzymes catalysing active transport, was adapted from a model originally discussed by Jardetzky⁴⁶, generalized to other systems by Singer⁴⁷, and described in Klingenberg's article⁴⁴.

This proposal is only barely consistent with what is known about the structure of $(Na^+ + K^+)$ -ATPase. Purified canine renal $(Na^+ + K^+)$ -ATPase contains two polypeptide chains, one a larger (α , molecular weight (MW) 120,000 $\pm$ 10,000), relatively hydrophobic protein, and the other a smaller (β , MW 55,000 $\pm$ 5,000) sialoglycoprotein^{43,48}. It is the larger polypeptide which is the sibling of the polypeptides from Ca^{2+} -ATPase and $(H^+ + K^+)$ -ATPase and it is generally assumed that this chain has the catalytic role. Both these polypeptides, however, are part of a specific complex⁴³ because they can be cross-linked to form a unique $\alpha\beta$ heterodimer^{48,49}. The two chains are present in the native enzyme in equimolar ratio⁴³ so that the minimum asymmetric unit is $\alpha\beta$. In certain conditions, it is possible to form a unique, covalently cross-linked α_2 dimer^{36,49,50}. Discrete complexes with apparent molecular weights of 380,000 (ref. 51) and 280,000 (ref. 52) and which are capable of regaining activity have been produced by detergent treatment. Electron micrographs have also been interpreted as indicating that the native enzyme is a multimer of $\alpha\beta$ units⁵³. The presently accepted, although far from proved, structure of native $(Na^+ + K^+)$ -ATPase is shown in Fig. 1. If this structure is accurate, then Ca^{2+} -ATPase and $(H^+ + K^+)$ -ATPase must have the same appearance except that the β chain would be missing.

Klingenberg⁴⁴ has proposed that the present experimental evidence is sufficient to conclude that the structure shown in Fig. 1 is correct for $(Na^+ + K^+)$ -ATPase. In addition, he assembles other experimental results which suggest that oligomeric structures are common in membrane-bound proteins. Although this may be true, it is far from proved, and it would be useful here to review briefly the serious shortcomings of the procedures which were used to reach these conclusions.

Oligomeric associations

In the case of covalent cross-linking, the geometry of the membrane itself confuses the result. In all membranes or membrane fragments, the distances between neighbouring protein molecules are much shorter than the distances between molecules in even the most concentrated isotropic solutions of soluble proteins. As a result, collisions between unassociated molecules are much more frequent, increasing the likelihood of artefactual cross-linking. Although the overall cross-linking reactions are often pseudo-first order in their kinetics³⁶, this is consistent with either a preexisting complex or an initial rate-limiting covalent reaction followed by a very rapid collision of monomers. An additional problem caused by the two-dimensional properties of the bilayer is that on every monomer of a certain sequence, each of the specific amino acid residues which are the targets for a given cross-linking reagent will lie at the same distance from the centre of the bilayer as its twin on another monomer. When bis-functional reagents are used, this fact greatly increases the likelihood that the twins will be coupled on collision. Once a covalent, symmetrical pair is formed, however, from two previously unacquainted monomers, neither has the reactive position available for further polymerization and a unique covalent dimer will be observed with no evidence of higher oligomers. A striking example of this principle is the adventitious dimer of acetylcholine receptors which is found in preparations made in the absence of reducing reagents but which serves no biological purpose⁵⁴. Finally, cross-linking of membrane-bound proteins is rendered more ambiguous by the presence of the randomly oriented clusters of proteins often seen in electron micrographs⁵³.

The oligomeric assemblies of membrane-bound proteins which have been observed in the ultracentrifuge when these molecules are dissolved in various detergents are also cited⁴⁴ as evidence for the existence of similar, native complexes. It is not the case, however, that only one, unambiguous complex is always observed under these circumstances. Stable tetramers, dimers and monomers of Ca^{2+} -ATPase have been separated and identified⁵⁵. $(\text{Na}^+ + \text{K}^+)\text{ATPase}$, on the other hand, is supposed to exist as a dimer in detergent solution, compatible with the fantasy shown in Fig. 1. However the molecular weights, determined for this presumably discrete complex by ultracentrifugation in two separate laboratories^{51,52}, differ by 30% from each other. Furthermore, in neither case did the protein show $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ activity when dissolved in the concentration of detergent required to disperse it. Finally, the very substantial corrections which are necessary to account for the effects of bound detergent, carbohydrate and lipid on the hydrodynamic properties of the particle⁵¹, although theoretically pleasing, have yet to be independently verified. The cautious reader should note the errors made with aldolase⁵⁶ and aspartate transcarbamoylase⁵⁷ with the assistance of an even less complicated theoretical development in much simpler circumstances.

A clear indication of the uncertainty of the present position is that Klingenberg⁴⁴ has concluded, from experimental results of the type just discussed, that $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ is a dimer of asymmetric units, whereas Ca^{2+} -ATPase is a tetramer. If this

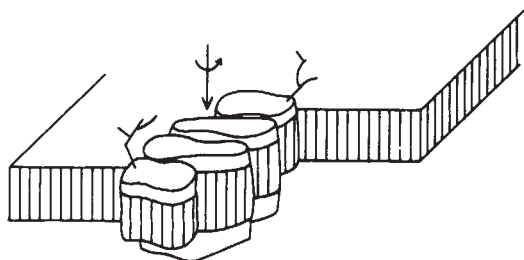


Fig. 1 Drawing of the imagined molecular structure of $(\text{Na}^+ + \text{K}^+)\text{ATPase}$. The subunits are drawn so that their volumes correspond to molecular weights of 120,000 and 55,000 using as a scale the width of the bilayer (4.0 nm).

were true, $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ would have a two-fold symmetry axis and Ca^{2+} -ATPase, a four-fold one. As it is impossible to evolve a protein with a four-fold axis from one with a two-fold axis, or vice versa, the common ancestor of these proteins would have been a monomer. If this were the case, the oligomeric structures of these proteins would be irrelevant to the process of active transport. A more likely resolution of this confusion, however, is that both $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ and Ca^{2+} ATPase have the same oligomeric structure but the true aggregation state of neither has been established.

Half-of-sites behaviour

Unfortunately, the theoretical proposals for the mechanism of active transport and the experiments which investigated the quaternary structures of membrane-bound proteins were being published at a time when enthusiasm for half-of-sites mechanisms of enzyme catalysis had reached its peak⁵⁸. It was immediately apparent that active transport was a strong candidate for half-of-sites behaviour, and many results appeared which supported the possibility. There are several ligands which bind to $(\text{Na}^+ + \text{K}^+)\text{ATPase}$: ouabain, vanadate, ATP and the phosphate of the covalent intermediate. Whenever comparisons are made with the same enzyme preparation, the molar concentrations of each of these sites are the same⁵⁹⁻⁶³. As such, these ligands define the concentration of functional $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ units or active sites. There have been many estimates of the grams of protein per mol of sites in purified $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ —these usually lie between 220,000 and 280,000 g per mol of sites⁵⁹⁻⁶³. If one decides, on the basis of electrophoretic mobility, that the molecular weights of α and β are 90,000 and 35,000, respectively, it can be concluded that there is about one site for the ligands per $(\alpha\beta)_2$ or 'half-of-sites'.

This conclusion had three major weaknesses. Most of the protein concentrations were based on the Lowry method which does not yield accurate, absolute estimates of protein concentration unless extinction coefficients are independently determined⁶⁴. Second, any estimate of grams per mol of sites necessarily represents only an upper limit of the true number because some inactive protein is always present. Therefore, the lower numbers are the more important estimates. Finally, the molecular weights quoted for the chains are based only on gel electrophoresis, which gives minimum estimates of the true molecular weight of a hydrophobic protein⁶⁵. It has been pointed out that SDS-polyacrylamide gel electrophoresis underestimates the molecular weight of the subunit of Ca^{2+} -ATPase as a determination of this quantity by sedimentation equilibrium indicates that the true molecular weight of this polypeptide chain is 120,000 (ref. 66).

New estimates of ouabain and ATP site concentrations for $(\text{Na}^+ + \text{K}^+)\text{ATPase}$, based on protein determination by quantitative amino acid analysis, yield values of $180,000 \pm 20,000$ g per mol of sites^{22,67}. Similar values have also been obtained recently for the concentration of phosphorylation sites⁶⁸. Finally, it has been established that there are 0.8–0.9 phosphorylation sites^{10,69} and ATP-binding sites⁶⁹ present on each polypeptide chain in purified Ca^{2+} -ATPase. Clearly there is sufficient reason to doubt that active transport enzymes ever display half-of-sites behaviour.

In this context, it is illuminating to consider the present state of the half-of-sites or negative cooperativity problem in general. With the exception of hexokinase⁷⁰, which for symmetry reasons is irrelevant to the problem of membrane-bound proteins, the number of established cases of half-of-sites behaviour is rapidly dwindling. There are many reasons why mistakes were made, even with soluble proteins, and each is more likely to occur when a membrane-bound enzyme is examined. The most common problem is heterogeneity of enzyme due to the presence of contaminating, inactive protein⁷¹ of partially denatured enzyme of lower affinity^{72,73} or another protein which alters the properties of some of the enzyme⁷⁴. It is remarkable how such situations which should lead to various degrees of heterogeneous behaviour always produce half-of-sites behaviour; the concen-

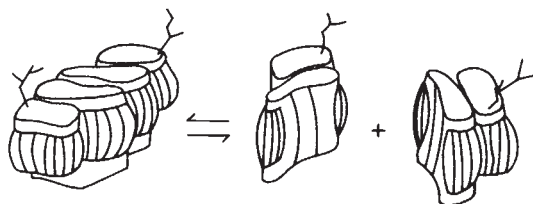


Fig. 2 Hypothetical equilibrium between two monomers and a dimer of $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ asymmetric units dissolved in detergent. The detergent replaces the bilayer as a collar around the protein.

tration of sites is always half that expected. A corollary of these considerations would be the consequences which would result if some of the enzyme had been completely denatured during the detergent treatments necessary to isolate membrane-bound proteins. The partial or complete demise of an enzyme almost never alters its behaviour on polyacrylamide gels and would pass unnoticed in the ultracentrifuge. Another problem encountered with membrane-bound proteins is the presence of vesicles and multibilayers which prevent access of ligands to their otherwise normal sites. For example, the concentration of cardiac glycoside sites on $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ increases 1.6-fold when a membrane-bound, vesicular preparation is dissolved in detergent⁶⁷ even though the polyacrylamide gels of the two preparations are indistinguishable.

Finally, the shortcomings of the analytical procedures themselves must be considered. Any practising protein chemist realizes that the determination of the actual protein concentration in an experimental sample is extremely difficult²². A direct approach to the determination of the concentration of sites in a given preparation is often plagued by the problems of unstable intermediates, loss of protein, or large concentrations of unbound ligand—difficulties which can be overcome, however, by using more imaginative procedures⁷⁵. Often, the actual molecular weight of the subunits is doubtful. This quantity can only be established unambiguously from the sequence⁷⁶, but few sequences of membrane-bound proteins are available. Those

which are show that SDS-polyacrylamide gel electrophoresis significantly underestimates their true polypeptide length⁷⁷.

Almost all these difficulties result in an apparent decrease in the concentration of sites relative to the concentration of the polypeptide and may have contributed to the erroneous conclusion that half-of-sites behaviour is shown by active transport enzymes.

Present uncertainty

Thus it is unknown whether the boundary between two α subunits is involved in the process of active transport. The hyperbolic, non-interacting binding curves which are generally observed for the various ligands imply that the sites on each chain are well separated from each other and certainly not located within the same active site. Although dispersed forms of $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ and $\text{Ca}^{2+}\text{-ATPase}$ which seem to be multi-meric have been identified, their existence does not necessarily imply that such arrangements are critical for function, something half-of-sites could have settled. Reports which propose that $\text{Ca}^{2+}\text{-ATPase}$ can function as a monomer further complicate the issue⁷⁸⁻⁸⁰. If this were true, it would eliminate the hypothesis outlined above, as well as any consideration of 'half-of-sites', in the case of active transport. There are several distinct difficulties, however, with these experiments. First, estimations of the aggregation state of the enzyme were based only on hydrodynamic parameters which are difficult to interpret in systems containing protein, lipid and detergent. Furthermore, no attempt was made to determine the quaternary structure of the enzyme at the same time that turnover was being observed. If an equilibrium such as the one depicted in Fig. 2 can occur, then the conditions of the enzyme assay may promote dimerization and reactivation even though an inactive monomer is present in the ultracentrifuge. In conclusion, it is unknown whether the functional unit of an active transport protein is monomeric or dimeric, and if a dimer is involved whether it can participate in the way described above and yet not display half-of-sites behaviour.

J. K. was supported by a NIH Research Career Development Award.

- Kyte J. *J. biol. Chem.* **246**, 4157-4165 (1971).
- MacLennan, D. H. *J. biol. Chem.* **245**, 4508-4518 (1970).
- Niggli, V., Penniston, J. T. & Carafoli, E. *J. biol. Chem.* **254**, 9955-9958 (1979).
- Sachs, G. *et al. J. biol. Chem.* **251**, 7690-7698 (1976).
- Sen, A. K. & Post, R. L. *J. biol. Chem.* **239**, 345-352 (1964).
- Hasselbach, W. *Biochim. biophys. Acta* **515**, 23-53 (1978).
- Scarborough, G. A. *Biochemistry* **19**, 2925-2931 (1980).
- Dufour, J. P. & Goffeau, A. *J. biol. Chem.* **253**, 7026-7032 (1978).
- Uesugi, S. *et al. J. biol. Chem.* **246**, 531-543 (1971).
- MacLennan, D. H., Seeman, P., Iles, G. H. & Yip, C. C. *J. biol. Chem.* **246**, 2702-2710 (1971).
- Saccomani, G., Shah, G., Spennay, J. G. & Sachs, G. *J. biol. Chem.* **250**, 4802-4809 (1975).
- Dame, J. B. & Scarborough, G. A. *Biochemistry* **19**, 2931-2937 (1980).
- Bastide, F., Meissner, G., Fleischer, S. & Post, R. L. *J. biol. Chem.* **248**, 8385-8391 (1973).
- Karlsh, S. J. D., Jørgensen, P. L. & Gitler, C. *Nature* **269**, 715-717 (1977).
- Farley, R. A., Goldman, D. W. & Bayley, H. *J. biol. Chem.* **255**, 860-864 (1980).
- Allen, G., Trinnamen, B. J. & Green, N. M. *Biochem. J.* **187**, 591-616 (1980).
- Castro, J. & Farley, R. A. *J. biol. Chem.* **254**, 2221-2228 (1979).
- Post, R. L. & Kume, S. *J. biol. Chem.* **248**, 6993-7000 (1973).
- Albers, R. W., Fahn, S. & Koval, G. *J. Proc. natn. Acad. Sci. U.S.A.* **50**, 474-481 (1963).
- Kanazawa, T., Saito, M. & Tonomura, Y. *J. Biochem.* **67**, 693-711 (1970).
- Karlsh, S. J. D., Yates, D. W. & Glynn, I. M. *Biochim. biophys. Acta* **525**, 252-264 (1978).
- Moczydlowski, E. G. & Fortes, P. A. G. *J. biol. Chem.* **256**, 2346-2356; 2357-2366 (1981).
- Karlsh, S. J. D. & Yates, D. W. *Biochim. biophys. Acta* **527**, 115-130 (1978).
- Jørgensen, P. L. *Biochim. biophys. Acta* **401**, 399-415 (1975).
- Koepsell, H. *J. Membrane Biol.* **48**, 69-94 (1979).
- Fukushima, Y. & Tonomura, Y. *J. Biochem.* **74**, 135-142 (1973).
- Fahn, S., Hurley, M. R., Koval, G. J. & Albers, R. W. *J. biol. Chem.* **241**, 1890-1895 (1966).
- Fahn, S., Koval, G. J. & Albers, R. W. *J. biol. Chem.* **243**, 1993-2002 (1968).
- Mårdh, S. & Post, R. L. *J. biol. Chem.* **252**, 633-638 (1977).
- Beaugé, L. A. & Glynn, I. M. *J. Physiol., Lond.* **289**, 17-31 (1979).
- Dupont, Y. & Leigh, J. B. *Nature* **273**, 396-398 (1978).
- Takisawa, H. & Tonomura, Y. *J. Biochem.* **86**, 425-441 (1979).
- Wallmark, B., Stewart, H. B., Rabon, E., Saccomani, G. & Sachs, G. *J. biol. Chem.* **255**, 5313-5319 (1980).
- Neet, K. E. & Green, N. M. *Archs Biochem. Biophys.* **178**, 588-597 (1977).
- Neufeld, A. H. & Levy, H. M. *J. biol. Chem.* **244**, 6493-6497 (1969).
- Kyte, J. *J. biol. Chem.* **250**, 7443-7449 (1975).
- Bretscher, M. S. *J. molec. Biol.* **59**, 351-357 (1971).
- Singer, S. J. & Nicolson, G. N. *Science* **175**, 720-731 (1972).
- Khorana, H. G. *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 5046-5050 (1979).
- Jost, P. C., Griffith, O. H., Capaldi, R. A. & Vanderkooi, G. *Proc. natn. Acad. Sci. U.S.A.* **70**, 480-484 (1973).
- Monod, J., Wyman, J. & Changeux, J. P. *J. molec. Biol.* **12**, 88-118 (1965).
- Cullis, A. F., Muirhead, H., Perutz, M. F., Rossmann, M. G. & North, A. C. T. *Proc. R. Soc. A* **268**, 161-187 (1962).
- Craig, W. S. & Kyte, J. *J. biol. Chem.* **255**, 6262-6269 (1980).
- Klingenberg, M. *Nature* **290**, 449-454 (1981).
- Kyte, J. *J. biol. Chem.* **249**, 3652-3660 (1974).
- Jardetzky, O. *Nature* **211**, 969-970 (1966).
- Singer, S. J. *A. Rev. Biochem.* **43**, 805-833 (1974).
- Kyte, J. *J. biol. Chem.* **247**, 7642-7649 (1972).
- Liang, S. M. & Winter, C. G. *J. biol. Chem.* **252**, 8278-8284 (1977).
- Giotta, G. J. *J. biol. Chem.* **251**, 1247-1252 (1976).
- Hastings, D. F. & Reynolds, J. A. *Biochemistry* **18**, 817-821 (1979).
- Esmann, M., Skou, J. C. & Christiansen, C. *Biochim. biophys. Acta* **567**, 410-420 (1979).
- Deguchi, N., Jørgensen, P. L. & Maunsbach, A. B. *J. Cell Biol.* **75**, 619-634 (1977).
- Anholt, R., Lindstrom, J. & Montal, M. *Eur. J. Biochem.* **109**, 481-487 (1980).
- le Maire, M., Möller, J. V. & Tanford, C. *Biochemistry* **15**, 2336-2342 (1976).
- Schachman, H. K. & Edelstein, S. J. *Biochemistry* **5**, 2681-2705 (1966).
- Weber, K. *J. biol. Chem.* **243**, 543-546 (1968).
- Fersht, A. R. *Biochemistry* **14**, 5-12 (1975).
- Peterson, G. L., Ewing, R. D., Hootman, S. R. & Conte, F. P. *J. biol. Chem.* **253**, 4762-4770 (1978).
- Jørgensen, P. L. *Biochim. biophys. Acta* **356**, 53-67 (1974).
- Jørgensen, P. L. *Biochim. biophys. Acta* **466**, 97-108 (1977).
- Perrone, J. R., Hackney, J. F., Dixon, J. F. & Hokin, L. E. *J. biol. Chem.* **250**, 4178-4184 (1975).
- Cantley, L. C. Jr, Cantley, L. G. & Josephson, L. J. *J. biol. Chem.* **253**, 7361-7368 (1978).
- Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. *J. biol. Chem.* **193**, 265-275 (1951).
- deJong, W. W., Zweers, A. & Cohen, L. H. *Biochem. biophys. Res. Commun.* **82**, 532-539 (1978).
- Rizzolo, L. J., LeMaire, M., Reynolds, J. A. & Tanford, C. *Biochemistry* **15**, 3433-3437 (1976).
- Kyte, J. *J. biol. Chem.* **247**, 7634-7641 (1972).
- Askari, A., Huang, W. & Antieau, J. M. *Biochemistry* **19**, 1132-1140 (1980).
- Meissner, G. *Biochim. biophys. Acta* **298**, 906-926 (1973).
- Steitz, T. A., Fletcher, R. J., Anderson, W. F. & Anderson, C. M. *J. molec. Biol.* **104**, 197-222 (1976).
- Monteilhet, C. & Blow, D. M. *J. molec. Biol.* **122**, 407-417 (1978).
- Gennis, L. S. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3928-3932 (1976).
- Scheek, R. M., Berden, J. A., Hooghiemstra, R. & Slater, E. C. *Biochim. biophys. Acta* **569**, 124-134 (1979).
- Matturo, J. M. & Hollenberg, M. D. *Proc. natn. Acad. Sci. U.S.A.* **75**, 3070-3074 (1978).
- Cocivera, M., McManaman, J. & Wilson, I. B. *Biochemistry* **19**, 2901-2907 (1980).
- Rosenbusch, J. P. & Weber, K. *J. biol. Chem.* **246**, 1644-1657 (1971).
- Bonitz, S. G., Coruzzi, G., Thalenfeld, B. E., Tzagoloff, A. & Macino, G. *J. biol. Chem.* **255**, 11927-11941 (1980).
- Jørgensen, K. E., Lind, K. E., Røsgaard-Petersen, H. & Möller, J. V. *Biochem. J.* **169**, 489-498 (1978).
- Dean, W. L. & Tanford, C. *Biochemistry* **17**, 1683-1690 (1978).
- Møller, J. V., Lind, K. E. & Andersen, J. P. *J. biol. Chem.* **255**, 1912-1920 (1980).

ARTICLES

Detecting CO₂-induced climatic change

T. M. L. Wigley & P. D. Jones

Climatic Research Unit, University of East Anglia, Norwich NR4 7TJ, UK

Although it is widely believed that increasing atmospheric CO₂ levels will cause noticeable global warming, the effects are not yet detectable, possibly because of the 'noise' of natural climatic variability. An examination of the spatial and seasonal distribution of signal-to-noise ratio shows that the highest values occur in summer and annual mean surface temperatures averaged over the Northern Hemisphere or over mid-latitudes. The spatial and seasonal characteristics of the early twentieth century warming were similar to those expected from increasing CO₂ based on an equilibrium response model. This similarity may hinder the early detection of CO₂ effects on climate.

OVER the past 100 yr there has been a steady increase of atmospheric carbon dioxide levels¹ with the concentration rising from ~315 p.p.m. in 1958 to 338 p.p.m. in 1980^{2,3}. Most of this increase can be attributed to the ever-increasing use of fossil fuels. Carbon budget studies indicate that atmospheric CO₂ may reach 600 p.p.m., approximately double the pre-industrial level, some time between AD 2030 and 2080^{4,5}. A doubling of CO₂ levels may well cause a major climatic change: however, the increase so far is probably too small for attendant climatic change to be detected above the 'noise' of natural climatic variability^{6,7}.

It has therefore been suggested⁸ that attention should be directed to the identification of a 'precursor signal' of the effects of CO₂ on climate. This requires a knowledge of the signal and a knowledge of the noise. The signal can be estimated only by numerical modelling experiments; the noise, a result of the considerable natural variability of climate, can be estimated using recent instrumental data. Simple radiative-convective energy balance models⁹ indicate that a doubling of CO₂ could produce an increase in global mean annual surface temperature of 2–3 °C. More complex models^{10–12} give information about the spatial and seasonal patterns of climatic change, but there is still considerable uncertainty in these model simulations. Nevertheless, for any given modelled signal we can estimate, using recent data, the corresponding noise and so calculate a signal-to-noise ratio. The best climate parameters to monitor are those with the highest signal-to-noise ratio. These are not necessarily those with the highest signal.

Signal-to-noise ratio

One approach is to look for significant changes in the mean values of particular climate parameters. Climatic noise manifests itself through uncertainties in the means, which in turn can be quantified by the statistical confidence limits for the sample mean ($\bar{X}$) as an estimate of the population mean. For a sample of size N (yr) with variance $\hat{s}_N^2 (= \{\sum_{i=1}^N (X_i - \bar{X})^2\} / (N-1))$, the confidence limits' range, which is proportional to $\hat{s}_N / \sqrt{N}$, decreases roughly as $1/\sqrt{N}$. Averaging over longer time periods gives a more confident estimate of the mean and reduces the noise level. However, the reduction in noise is not strictly proportional to $1/\sqrt{N}$. Because most climate variables have a considerable amount of low-frequency variance, $\hat{s}_N$ increases with increasing sample size. Time-averaging reduces the noise, but not quite as rapidly as $1/\sqrt{N}$. To estimate the effect of low-frequency variance, we suppose climate time series to behave like first-order autoregressive processes. An additional factor

where r is the lag-1 autocorrelation, appears, and the noise level can be defined as

$$\eta = \sqrt{2}(\hat{s}_N / \sqrt{N})f(N, r) \quad (1)$$

The signal-to-noise ratio is simply (signal)/ η . The $\sqrt{2}$ term arises because, in comparing two means, both will be subject to noise. As $f(N, r) > 1$, autocorrelation, or any type of enhanced low-frequency variance, effectively increases the noise level, and its presence makes the statistical detection of a signal more difficult.

Madden and Ramanathan⁷ have calculated the reduction in noise in seasonal and annual mean surface air temperatures at 60°N due to time averaging, allowing for changes in $\hat{s}_N$. Their analysis suggests that, in the absence of other compensating effects, and ignoring lag effects due to the thermal inertia of the oceans, the CO₂ signal should be detectable either now or by about AD 2000, depending on the assumed character of the signal. They conclude that the effects at 60°N should be detectable most easily in summer.

Results for 60°N need not apply to other latitudes, and it may well be that a higher signal-to-noise ratio obtains at some other latitude band. Furthermore, because spatial averaging reduces the noise level in most climate parameters, averages over wider latitudinal ranges may increase the signal-to-noise ratio. We test these possibilities in this article.

Madden and Ramanathan used the results of Manabe and Wetherald¹⁰ and Ramanathan *et al.*¹³ to estimate the signal. In the latter work, the maximum signal is in high latitudes and in late spring-early summer. Here we use the general circulation model results of Manabe and Stouffer¹² in which the signal is greatest at high latitudes and in late autumn to winter. Although the signal should certainly differ from season to season, there is some doubt about the seasonal distribution⁷.

For our noise data we use gridded (5° latitude by 10° longitude) monthly surface air temperature data for the Northern Hemisphere over the period 1941–80 from ref. 14. The basic data are from World Weather Records, updated by NOAA in Monthly Climatic Data for the World, and supplemented with data from various National Meteorological Services, published records, and from the CLIMAT network. The number of individual stations varies from ~800 to 1,300. Data coverage becomes less complete as one moves back from 1940. Because of the sparsity of meteorological stations in ocean areas, the data grid has quite large gaps over the oceans. We have therefore used the land-area signal given by Manabe and Stouffer¹² (their Fig. 16c) to calculate signal-to-noise ratios.

Spatial distribution of signal-to-noise ratio

In Fig. 1 we show the signal-to-noise ratio (SN) as a function of latitude and month which is defined as

$$SN = \frac{1}{2} \Delta T_{MS} / \eta \quad (2)$$

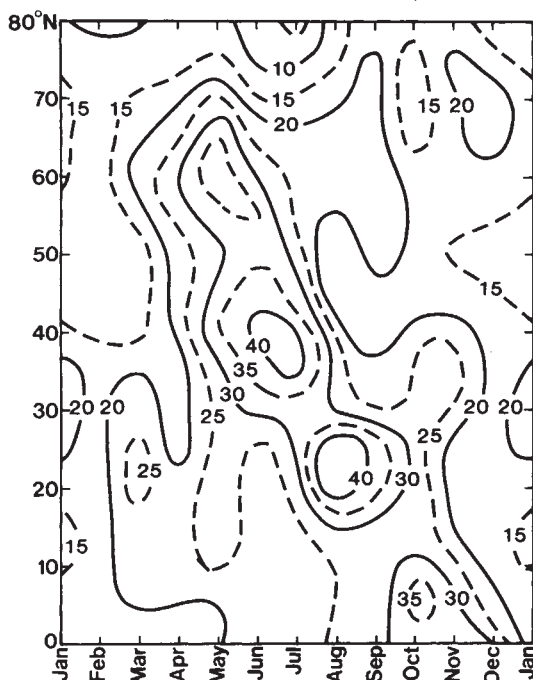


Fig. 1 Signal-to-noise ratio for predicted CO_2 -induced changes in surface air temperature as a function of latitude and month. The signal is based on the numerical modelling results of Manabe and Stouffer¹². The noise has been calculated from grid-point surface temperature data¹⁴: the value for month j at latitude L is the areally weighted average of grid points at $L-5$, L and $L+5$, and the noise level is proportional to the standard deviation of month- j values over the period 1941–80, corrected for autocorrelation effects.

where η is given by equation (1), and ΔT_{MS} is the temperature increase for a quadrupling of atmospheric CO_2 (the factor $1/2$ arises because ΔT is approximately proportional to $\ln q$ where $q = \text{CO}_2/\text{CO}_2^0$ (ref. 7)). We have allowed for the effect of low-frequency variance on time-averaging by including the term $f(N, r)$. $f(N, r)$ is generally near 1 for our data, but ranges up to a maximum near 2. Its inclusion makes little qualitative difference to our results. Although our noise levels have been calculated using the period 1941–80 ($N=40$), we have examined other time periods (and different values of N) and obtained very similar results. A value of $\text{SN} \geq 1.7$ corresponds to a statistically significant ΔT at the 0.05 level. The temperature changes corresponding to a doubling of CO_2 are highly significant for all months and all latitudes relative to the noise level.

Figure 1 shows that the most likely season and latitude zone in which to detect the effects of CO_2 is during the summer months in middle latitudes; in spite of the fact that the assumed maximum signal is in winter in high latitudes. The importance of summer in detecting CO_2 effects is in accord with Madden and Ramanathan⁷, but our results suggest that the signal-to-noise ratio around 40°N is higher than at 60°N .

Monthly values of zonally averaged temperatures may not give the highest values of SN. Averaging into seasons and over the whole year reduces the variance of climate parameters, and a further variance reduction can be achieved by spatial averaging. We therefore examined seasonal and annual data for four different regions, a low-latitude zone (5°S – 25°N inclusive), a mid-latitude zone (25° – 55°N), a high-latitude zone (55° – 85°N) and the whole Northern Hemisphere (0 – 85°N). SN values, defined as above, are given in Table 1. From these results, and similar analyses of different time periods (not shown here), the highest signal-to-noise ratios appear to be for summer mid-latitudes, spring and summer Northern Hemisphere averages, annual mid-latitudes and annual Northern Hemisphere averages. Models with a different seasonal distribution of signal (see ref. 13) would imply even stronger signal-to-noise ratios in summer.

Recent changes in Northern Hemisphere and mid-latitude, summer and annual average surface temperatures are shown in Fig. 2. Some of the short-term variability has been smoothed out by using a 1–2–1 binomial filter, but the noisiness of the record is apparent. In neither of the summer curves is there any apparent trend (in either direction) since the mid-1960s. A slight upward trend is noticeable in the annual curves after 1972, but more pronounced and longer-lasting trends have occurred before, noticeably before 1940. We therefore examined this early warming period more closely.

Comparison with an earlier warming period

Figure 2 shows the strong warming trend in Northern Hemisphere and mid-latitude, summer and annual average temperatures between 1910 and 1938. If we compare the coldest and warmest 20-yr periods this century (1901–20 and 1934–53 respectively), then all seasons and latitude bands show a warming. Statistically, the most significant changes occurred in Northern Hemisphere averages for spring, summer, autumn and annual means, and in mid-latitude averages for summer and annual means, in striking accord with the latitude zones and seasons for which the CO_2 signal-to-noise ratio is highest (Table 1).

Although we can be reasonably sure that the early twentieth century warming was not due to increasing CO_2 , it is instructive to examine this possibility more closely. Two approaches may be used to test the hypothesis that CO_2 was a causal factor; the magnitude of warming can be compared with that expected from the known increase in atmospheric CO_2 , or the spatial and seasonal patterns of warming can be compared with those expected on the basis of CO_2 modelling experiments. The latter comparison is shown in Fig. 3 where the CO_2 signal¹² is plotted against the temperature change between 1901–20 and 1934–53 using the same seasonal and spatial grouping as before. The comparison is sufficiently good that it could be taken to support CO_2 as a cause for the early twentieth century warming. There are some interesting differences, however. Compared with the annual Northern Hemisphere temperature change, the observed low-latitude changes were too low, especially in summer. In mid-latitudes the seasonal character of the observed changes (winter minimum, summer maximum) contrasts sharply with that expected on the basis of Manabe and Stouffer's model, but

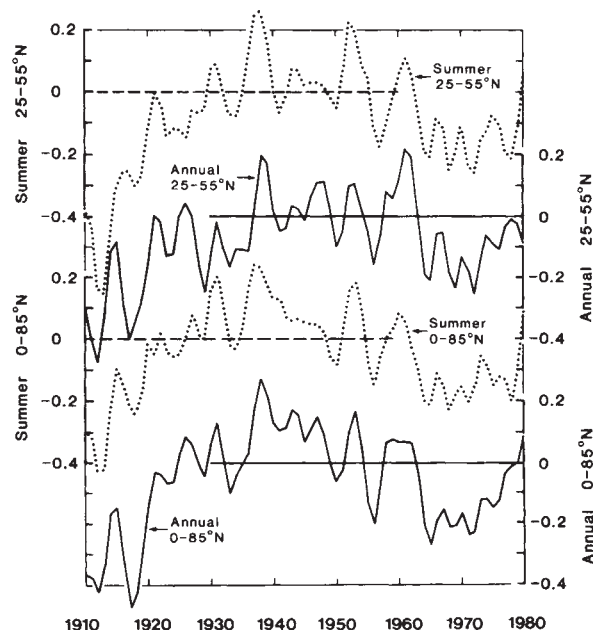


Fig. 2 Variations in Northern Hemisphere and mid-latitude (25 – 55°N) summer and annual average temperatures over the past 70 yr from ref. 14. The values are anomalies relative to the 1946–60 mean and have been filtered using a three-term binomial filter (except for the 1980 value).

accords better with Ramanathan *et al.*¹³. Given our lack of understanding of the details of both natural and CO₂-induced climatic change, these exceptions alone are probably insufficient to reject CO₂ as a possible cause of the early twentieth century warming.

However, the warming between 1901–20 and 1934–53 is, based on the most recent numerical model results, much larger than would be expected to result from the small increase in atmospheric CO₂ which occurred over this period. Additionally, between about 1940 and 1970 there was a significant cooling (Fig. 2) and this is incompatible with the CO₂ hypothesis.

The early twentieth century warming is a low-frequency component of the noise which acts to obscure any influence of increasing atmospheric CO₂ (in so far as any climatic change which is not attributable to CO₂ can be considered as noise). Because we have information about only one manifestation of this low-frequency component, we know nothing about its statistical properties. Nevertheless, this single example illustrates the similarity between natural and predicted CO₂-induced temperature changes; a similarity which might imply that analysis of spatial and/or seasonal details of future variations in surface temperature will be of little help in distinguishing natural changes from the effects of CO₂.

A second difficulty in detecting CO₂ effects arises because of lags in the response to increasing CO₂ which result from thermal inertia associated with oceanic mixing¹⁵. Most models used in studying possible CO₂ effects (including that of Manabe and Stouffer¹²) have considered the equilibrium response to a step-function increase in CO₂. Transient response models¹⁵ indicate that oceanic influences might cause the climate response to lag 5–20 yr behind the CO₂ forcing. Such a lag, in itself, is a significant factor in determining when the effects of CO₂ might first be detected. However, an equally important result is that the spatial pattern of the transient response may differ considerably from that expected on the basis of equilibrium models¹⁵. We have shown that the character of natural low-frequency changes in surface temperature can be similar to the equilibrium CO₂ response. If transient response patterns do differ from equilibrium response patterns, this may make it easier to distinguish CO₂ effects from natural changes in climate.

Conclusions

We have examined the signal-to-noise ratio for surface temperature averaged over periods of 1 month to 1 yr and over latitude zones of widths from about 15° to 85°. The quantitative details of our results, especially those in Fig. 1, are model dependent. Our general conclusions are, however, relatively insensitive to the choice of the model used to define the CO₂ signal, because they depend strongly on the spatial and seasonal distribution of the natural variability of surface temperatures. For example, our conclusion that signal-to-noise ratio is highest for mid-latitude or Northern Hemisphere summer or annual means arises primarily because of the relatively low natural variability of summer temperatures. We can also conclude that the signal-to-noise ratio in low latitudes is of comparable magnitude to that in high latitudes (see Table 1), primarily

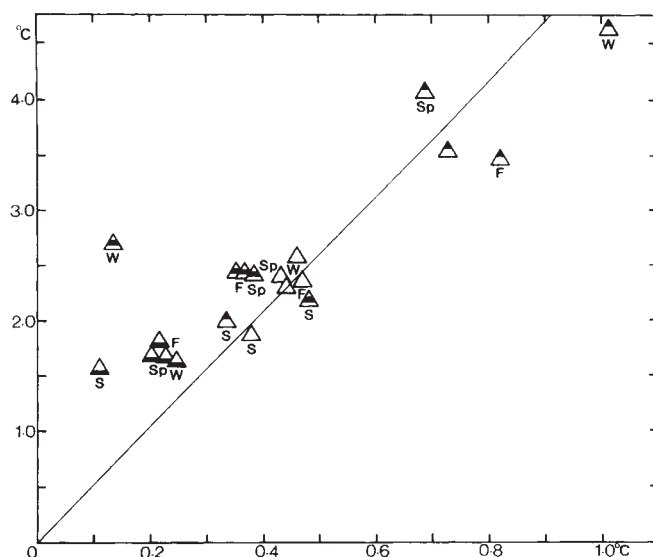


Fig. 3 Comparison of observed temperature changes between 1901–20 and 1934–53 with changes expected to result from increasing atmospheric CO₂ levels. The abscissa is the observed temperature change and the ordinate is the CO₂ signal for a doubling of CO₂ based on the land area results of Manabe and Stouffer¹². Different latitude bands are distinguished by the position of the band in the triangle. Δ , Hemispheric averages. Seasons: W, winter; Sp, spring; F, autumn or fall; no letter, annual average.

because of the low variability of low-latitude surface temperatures. This conclusion would only be invalidated if the ice-albedo feedback at high latitudes were considerably underestimated by currently available climate models: this is unlikely¹⁵. The uncertainty in the spatial patterns of climate response to increasing atmospheric CO₂ places considerable emphasis on monitoring large-scale averages for precursor signals of the effects of CO₂.

Because of the demonstrated similarity between observed natural changes in surface temperature and the changes predicted by Manabe and Stouffer's equilibrium model¹², it is important to investigate in more detail the differences in response between equilibrium and transient response models. If the differences are large, this would cast some doubt on the value of both equilibrium model results¹² and instrumental analogues¹⁶ in generating spatially detailed scenarios for a high-CO₂ warmer world. If the differences are small, the opposite conclusion could be drawn.

The change in surface temperature between 1901–20 and 1934–53 was similar in magnitude to that expected to result from a CO₂ increase of ~14% based on Manabe and Stouffer's model¹². Between 1970 and 1980 atmospheric CO₂ concentration increased by ~4%. A further 10% increase will almost certainly occur before the year 2000, probably by 1990. We might, therefore, anticipate changes in climate of a similar magnitude to those experienced in the first half of this century to occur over the next 10–20 yr. As most of the early twentieth century warming occurred between the late 1910s and the late 1930s (Fig. 2), the rate of future CO₂ induced changes this century is unlikely to be faster than natural changes which we have recently experienced. If current model simulations are correct, then the rate of change of surface temperatures over the next 20 yr is, therefore, not likely to be useful in distinguishing CO₂ effects from natural climatic change.

Because of difficulties in using surface temperature data to detect the effects of increasing atmospheric CO₂, it is important to look at other climate parameters. There is a strong statistical and causal link between circulation changes and changes in surface temperature¹⁷, so that surface climate variables other than temperature may not help significantly in trying to distinguish CO₂-induced changes from natural changes in climate.

Table 1 Signal-to-noise ratio for different latitude bands and seasons

Latitude range	Winter	Spring	Summer	Autumn	Annual
Low (5° S–25° N)	23 (1.6)	28 (1.7)	36 (1.6)	36 (1.8)	31 (1.7)
Mid (25° N–55° N)	26 (2.7)	36 (2.4)	49 (2.2)	40 (2.4)	53 (2.4)
High (55° N–85° N)	23 (4.6)	31 (4.1)	25 (2.0)	28 (3.5)	33 (3.5)
Northern Hemisphere (0–85° N)	32 (2.6)	44 (2.4)	43 (1.9)	35 (2.4)	44 (2.3)

The signal (°C) corresponding to a doubling of CO₂ (that is $\frac{1}{2}\Delta T_{MS}$) is given in parentheses; based on Manabe and Stouffer¹².

Detailed spatial studies are hampered by the uncertainties in the spatial patterns of CO₂ effects. The predicted cooling of stratospheric temperatures might be useful, but our knowledge of the natural variability of stratospheric temperatures is meagre. Furthermore, other anthropogenic effects might also cause stratospheric cooling¹⁸. The same problems apply to monitoring changes in lapse rate¹⁹. An interesting possibility might be to monitor satellite radiation budget data as suggested by Madden and Ramanathan⁷.

The noise level in climate data may be reduced by eliminating variations with known causes. For example, the Southern Oscillation^{20,21} accounts for a significant part of the climate variance in some parts of the world, and this component might usefully be factored out by regression techniques. Alternatively, if there were accurate measurements of atmospheric turbidity and/or solar variability to which aspects of climatic change could be unequivocally ascribed, then the effects of these parameters

could be removed, either by using numerical models or, less satisfactorily, by using empirically-based statistical relationships. Further improvements in numerical models and in our understanding of the causes of recent climatic change are required before this approach can give convincing results.

The effects of CO₂ may not be detectable until around the turn of the century. By this time, atmospheric CO₂ concentration will probably have become sufficiently high (and we will be committed to further increases) that a climatic change significantly larger than any which has occurred in the past century could be unavoidable. To avert such a change it is possible that decisions will have to be made (for example, to reduce anthropogenic CO₂ emissions) some time before unequivocal observational 'proof' of the effects of CO₂ on climate is available.

This work was supported by the US Department of Energy, Carbon Dioxide Research and Assessment Program, under contract DE-ACO2-79EV10098.

Received 23 March; accepted 27 May 1981.

1. Callendar, G. S. *Tellus* **10**, 243 (1958).
2. Keeling, C. D. *et al.* *Tellus* **28**, 538 (1976).
3. Keeling, C. D. *et al.* in *Global Energy Futures and the Carbon Dioxide Problem 2* (US Government Printing Office, Washington, DC, 1981).
4. Rotty, R. in *Man's Impact on Climate* (eds Bach, W., Pankrath, J. & Kellogg, W. W.) 269 (Elsevier, Amsterdam, 1979).
5. Zimen, K. E. in *Man's Impact on Climate* (eds Bach, W., Pankrath, J. & Kellogg, W. W.) 129 (Elsevier, Amsterdam, 1979).
6. Mitchell, J. M. Jr in *Workshop on the Global Effects of Carbon Dioxide from Fossil Fuels*, (eds Elliot, W. P. & Machta, L.) (DOE/CONF-770385, 91, 1979).
7. Madden, R. A. & Ramanathan, V. *Science* **209**, 763 (1980).
8. US Department of Energy *Research Issues and Supporting Research of the National Program on Carbon Dioxide, Environment and Society* (DOE/EV-1029, UC-11, 128, 1981).
9. Manabe, S. & Wetherald, R. T. *J. atmos. Sci.* **24**, 241 (1967).
10. Manabe, S. & Wetherald, R. T. *J. atmos. Sci.* **32**, 3 (1975).
11. Manabe, S. & Wetherald, R. T. *J. atmos. Sci.* **37**, 99 (1980).
12. Manabe, S. & Stouffer, R. J. *J. geophys. Res.* **85**, 5529 (1980).
13. Ramanathan, V., Lian, M. S. & Cess, R. D. *J. geophys. Res.* **84**, 4949 (1979).
14. Jones, P. D., Wigley, T. M. L. & Kelly, P. M. *Mon. Weath. Rev.* (submitted).
15. Schneider, S. H. & Thompson, S. L. *J. geophys. Res.* (in the press).
16. Wigley, T. M. L., Jones, P. D. & Kelly, P. M. *Nature* **283**, 17 (1980).
17. van Loon, H. & Williams, J. *Mon. Weath. Rev.* **104**, 365 (1976).
18. Ramanathan, V. & Dickinson, R. E. *J. atmos. Sci.* **36**, 1084 (1979).
19. Angell, J. K. & Korshover, J. *Mon. Weath. Rev.* **106**, 755 (1978).
20. Troup, A. J. *Q. Jl R. met. Soc.* **91**, 490 (1965).
21. Wright, P. B. *The Southern Oscillation—Patterns and Mechanisms of the Teleconnections and the Persistence* (Hawaii Institute of Geophysics, University of Hawaii, HIG-77-13, 1977).

Long-period geomagnetic secular variations since 12,000 yr BP

K. M. Creer

Department of Geophysics, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JZ, UK

A comparison of secular variation records of the geomagnetic field inclination obtained from palaeomagnetic studies of European and North American post-Glacial and late-Glacial sediments suggests that geomagnetic westward drift has persisted through the past 12,000 yr. Between 12,000 and 2,500 yr BP the overall average drift rate was between 0.14 and 0.16 deg yr⁻¹. At ~2,500 yr BP it is tentatively suggested that the drift rate increased and rose to a maximum of nearly 0.5 deg yr⁻¹ before decreasing to historically observed values of ~0.25 deg yr⁻¹.

OUR present knowledge of the spectrum and spacial characteristics of the geomagnetic field's longer period secular variations is derived from observatory magnetometer records which cover only the past two centuries, supplemented by the occasional magnetic compass and dip circle readings logged by ancient mariners¹ thus limiting our understanding of how the field is generated. From these data, world maps contouring the various geomagnetic elements have been constructed for epochs through the past three to four centuries²⁻⁴ from spherical harmonic coefficients computed up to degree $n = 4$.

It is convenient to separate the observed field at any specific epoch into dipole and non-dipole components. The pattern of the non-dipole part of the field has been distributed around four positive and five negative foci throughout history. Geomagnetic secular variations are caused both by changes in intensity and orientation of the dipole and by westward drift and changes in intensity of the non-dipole part of the field. Through the past four centuries, some non-dipole foci have drifted to the west while others have stood still: some have maintained constant intensity while others have grown or decayed^{5,6}.

The geomagnetic field can be separated mathematically into drifting and standing parts⁷⁻⁹. The form of the drifting part of the vertical component of the non-dipole field is shown in Fig. 1 for the Northern and Southern Hemispheres using Yukutake and

Tachinaka's data⁸. Its contours exhibit an essentially quadrupolar symmetry: in each hemisphere there is one large positive and one large negative region, each of which contains its own individual structure which evolves with time. The mean rate of westward drift is ~0.25 deg yr⁻¹ which is similar to that determined by Barraclough² for the eccentric dipole into which the dipole and quadrupole parts of the field are sometimes combined. The direction of drift is not purely westward¹⁰; in particular, it seems that the general quadrupole has rotated clockwise at a rate of 0.25 ± 0.04 deg yr⁻¹ during the present century¹¹ about a pole located near the Aleutian Islands rather than about the geographical pole.

The dipole component is not axial: its equatorial part has drifted to the west at an average rate of 0.29 deg yr⁻¹ since the sixteenth century¹⁰, though much slower drift rates have been obtained for epochs during the present century and have been explained by suggesting that the 'standing' part of the field actually undergoes a very slow eastward drift^{10,12}.

The lifetime of non-dipole sources was once thought to be as short as a few hundred years⁶, but recently substantially longer lifetimes of ~7,000 yr have been suggested¹⁰. For direct proof of such long lifetimes and of the persistency of westward drift we must look to archaeomagnetic and palaeomagnetic rather than observatory sources of data.

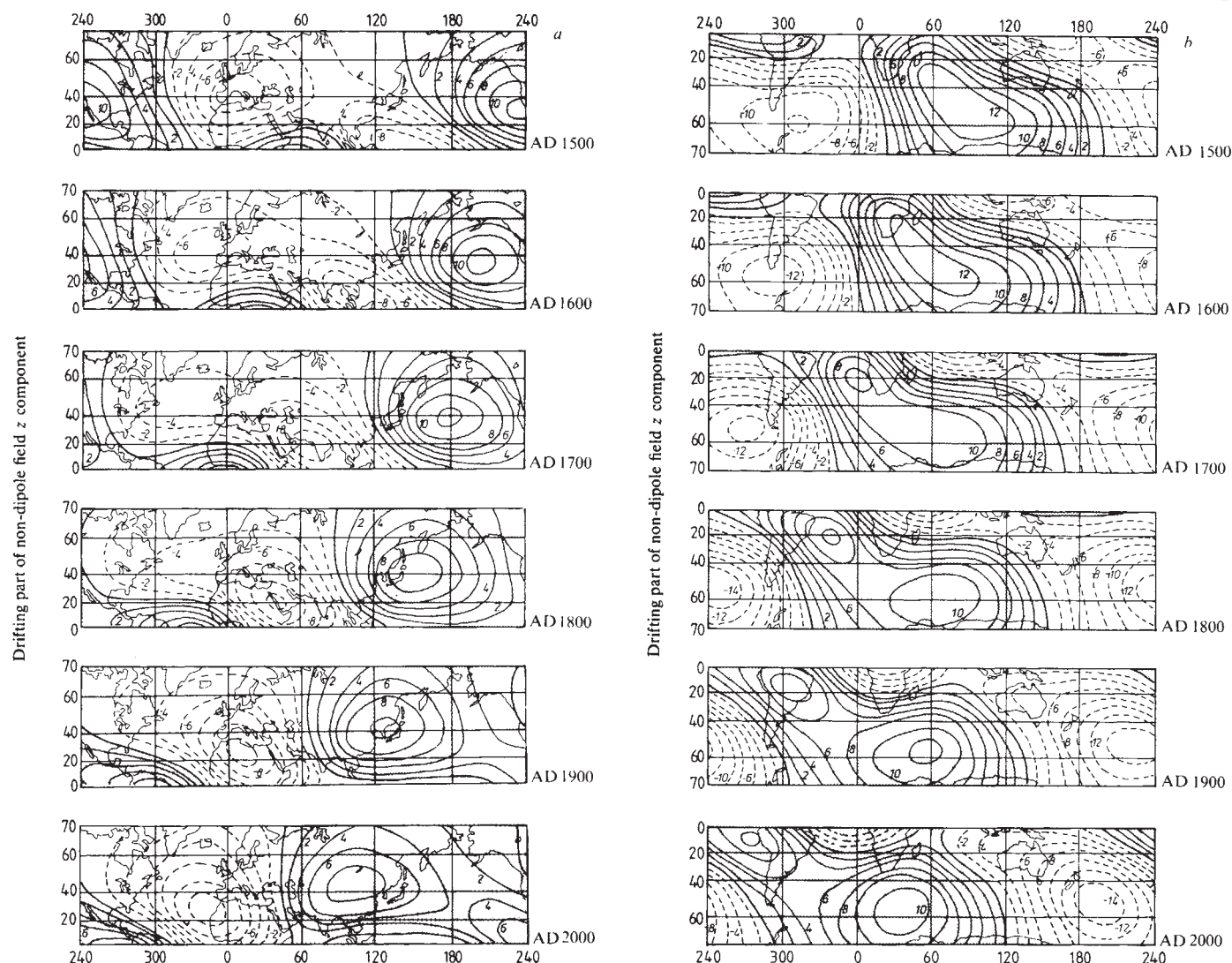


Fig. 1 The vertical component of the drifting part of the non-dipole field for six epochs from AD 1500 to AD 2000. *a*, Northern Hemisphere; *b*, Southern Hemisphere. Contour interval, 2,000 nT.

Thus, there are many gaps in our knowledge of the long term behaviour of the field during periods of settled polarity. The extended data sets becoming available from palaeomagnetic sources should help to resolve some of the outstanding problems.

Palaeomagnetic secular variation records

During the past decade significant developments have been made in techniques of taking high quality undisturbed cores of unconsolidated lake and marine sediments¹³ and it has become possible to measure their weak natural remanent magnetizations rapidly and accurately¹⁴. In favourable conditions of deposition this remanent magnetization has been found to carry a record of the secular variations of the direction of the ancient geomagnetic field at the site of deposition through the time the sediments were being accumulated. However, as might be expected, the quality of these natural records is much inferior to that we associate with instrumental records.

First, we must accept that the quality will be severely affected by any physical disturbance produced in the sediment, either naturally by slumping on the lake bottom due to, for example, to seismic activity or the action of turbidity currents, or mechanically during the coring process or during transport or during sub-sampling of cores. Another difficulty arises from the fact that corers often 'corkscrew' their way into the bottom sediments thus introducing a long wavelength trend and also dis-

continuities into the declination record. This is why more good quality inclination than declination records are available and why the present discussion is restricted to inclination secular variations.

Next, the natural remanent magnetization is not acquired instantaneously¹⁵: it is produced by a combination of different mechanisms such as post-depositional grain rotation, chemical changes and possibly bacterial action. Thus the recorded signal is always subjected to a variable degree of smoothing which explains why substantial variations in the amplitude of specific features identified along records from different places (see Figs 2 and 3) and even from different cores taken from the same lake are found.

Finally it is difficult to date recent sediments with the required degree of accuracy. Radiocarbon dates are often affected by the presence of recycled fossil carbon which may have been deposited into the sediment as fragments of fossil shell or plant remains ('old carbon' effect) or incorporated directly into organisms living at the time of deposition from calcium carbonate originating from pre-existing limestones dissolved in the lake water ('hard water' effect). Both these effects can produce radiocarbon ages which are too old by up to several thousand years so that we have to use indirect methods such as correlation of the 'palaeomagnetic' core with well-dated sections from the same locality using palynological or stratigraphical methods.

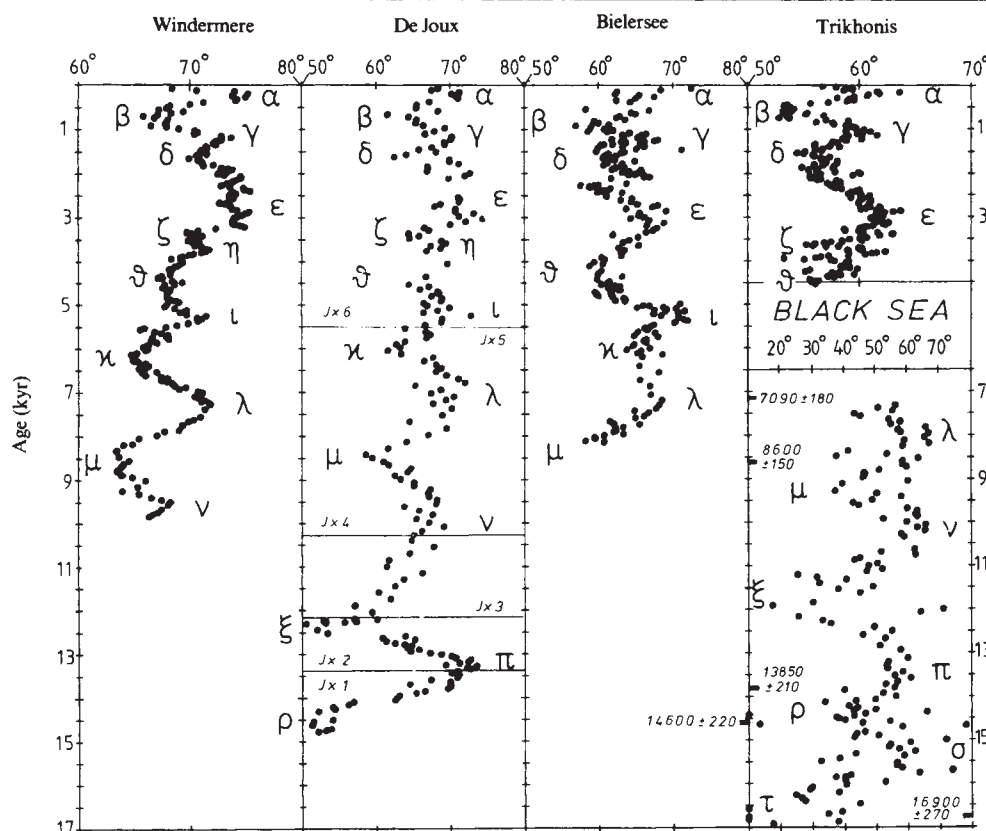


Fig. 2 Palaeomagnetic inclination logs for European sites. Time scale in uncalibrated radiocarbon years. Windermere¹⁸ and Trikhonis²⁰ curves are each stacked from several individual core records. Pollen zone boundaries shown on Lac de Joux log²¹. Radiocarbon ages shown along Black Sea log²².

European post-Glacial and late-Glacial data

Limnomagnetic research started with Mackereth's study¹⁶ of post-Glacial sediments from Lake Windermere. Subsequent studies in UK (Lake Windermere, Loch Lomond, Lake Geirionydd) have produced a detailed and reproducible pattern of secular variations in inclination and declination^{17,18} dated by over 30 radiocarbon age determinations. The characteristic features of the UK records, labelled α – ν , can also be recognized over a wide area of Europe from Poland¹⁹ to Greece²⁰. Records stacked from several cores from Lake Windermere (54°N 357°E)¹⁸ and from Lake Trikhonis (39°N 22°E)²⁰ are shown in Fig. 2. A record from Lac de Joux (47°N 5°E) extending back to about 14,000 yr BP has been dated by correlation with the well established biostratigraphy for the Joux valley²¹. It exhibits features θ – ρ particularly well. The younger features α – ι are exhibited more clearly for Switzerland in the Bielersee record, which as yet can be dated only by correlation of magnetic

declination and inclination features with the calibrated European records. Another inclination record extending from ~7,000 yr BP to ~25,000 yr BP (Fig. 2 goes to 17,000 yr BP) is provided by Black Sea core no.1474 (42°N 38°E) which was taken using a gravity piston corer. The time scale was constructed from seven radiocarbon age determinations²². Figure 2 thus shows the extent to which inclination secular variations can now be correlated across Europe back to ~15,000 yr BP.

North American post-Glacial and late-Glacial data

Inclination secular variation records from Lakes St Croix and Kylene, Minnesota (46°N 267°E)²³ and the Thunder Bay part of Lake Superior (48°N 271°E)²⁴ are shown in Fig. 3. Characteristic features which correlate from site to site are labelled α – π , although at this stage we should not identify them with the apparently similar features observed along the European records. At Lake St Croix parallel cores of organic rich gyttja were taken in sections by making successive 1.5-m drives using a piston corer. The cores were dated with eight radiocarbon age determinations²³ as indicated along the inclination record shown in Fig. 3. Thus the dating control is quite tight. At Lake Kylene, cores were taken using the same technique and dated by four radiocarbon age determinations. Note that there is a very large error of $\pm 1,500$ yr on the oldest age of 16,500 yr (ref. 22). Four cores were taken especially for palaeomagnetic study with an Alpine piston corer from Lake Superior²⁴. The age of the base of the post-Glacial was taken as 9,500 yr BP (ref. 25). Because sedimentation occurred in the Thunder Bay part of Lake Superior back to at least 11,000 yr BP (refs 25, 26), this age was taken as the age of the base of the core. The post-Glacial sediments were dated by nine radiocarbon age determinations some of which were affected by 'hard water' and 'old carbon' effects for which it is difficult to make a precise correction. The four oldest ages shown in Fig. 3 were accepted without correction and a 'zero-error' correction of 1,700 yr (following ref. 24) was applied to the three youngest ages. Thus this time scale is likely to be in error.

Though the pattern of variations provided by St Croix/Kylene record is not very well formed, the dating control is good. If the

Table 1 Ages of inclination features

Feature	Windermere (3°W)	St Croix (93°W)	Drift (yr)	Drift (deg yr ⁻¹)	Mean age (yr)
α	250	0	250	0.360	125
β	700	500	200	0.450	600
γ	1,150	900	250	0.360	1,025
δ	1,650	1,250	400	0.225	1,450
ϵ	3,000	2,200	800	0.112	2,600
ζ	3,600	2,700	900	0.100	3,150
η	4,000	3,600	400	0.225	3,800
θ	4,500	4,000	500	0.180	4,250
ι	5,200	4,600	600	0.150	4,900
κ	6,100	5,600	500	0.180	5,850
λ	7,300	6,700	600	0.150	7,000
μ	8,400	7,900	500	0.180	8,150
ν	10,100	9,400	800	0.124	9,750
ξ	12,400*	10,500†	1900	0.047	11,450
π	13,400*	12,600†	1800	0.112	13,000
ρ	14,500*				

Ages are given in radiocarbon yr BP.

* Lac de Joux not Windermere. † Kylene not St Croix.

Table 2 Quasi-periodicity of UK (Lake Windermere) and North American (Lake St Croix) inclination S.V.

Feature labels	Age difference (yr)	Mean age (yr)	Drift rate (deg yr ⁻¹)
Lake Windermere			
$\alpha-\gamma$	900	700	0.514
$\beta-\delta$	950	1,175	0.379
$\gamma-\epsilon$	1,850	2,075	0.195
$\delta-\theta$	2,850	3,075	0.126
$\epsilon-\iota$	2,200	4,100	0.163
$\theta-\kappa$	1,600	5,300	0.225
$\iota-\lambda$	2,100	6,250	0.171
$\kappa-\mu$	2,300	7,250	0.156
$\lambda-\nu$	2,800	8,700	0.129
$\mu-\xi^*$	4,000	10,400	0.090
$\nu-\pi^*$	3,300	11,750	0.109
$\xi-\rho^*$	2,100	13,450	0.171
Lake St Croix			
$\alpha-\gamma$	900	450	0.400
$\beta-\delta$	750	875	0.480
$\gamma-\epsilon$	1,300	1,550	0.277
$\delta-\theta$	2,750	2,625	0.131
$\epsilon-\iota$	2,400	3,400	0.150
$\theta-\kappa$	2,000	5,000	0.180
$\iota-\lambda$	1,900	5,650	0.189
$\kappa-\mu$	1,900	6,950	0.189
$\lambda-\nu$	2,500	8,050	0.144
$\mu-\xi$	2,600	9,200	0.138
$\nu-\pi$	2,100	10,450	0.171

All ages given in 'raw' radiocarbon years.

* Uses Lac de Joux, not Windermere, data.

ages of the labelled features obtained from this record are transferred to the rather well-formed Superior record a satisfactory inclination record for the Great Lakes region is obtained, the pattern of which is confirmed by recent results from Lake Huron²⁷.

Palaeowestward drift

We now consider whether the various characteristic features of the European and North American data labelled in Figs 2 and 3 can be correlated across the North Atlantic, remembering that the Greek labels were initially attached quite independently to both sets of records. In fact a rather striking overall similarity between the patterns, especially for features $\epsilon-\xi$ will be noted.

Considering the behaviour of the drifting part of the historic field shown in Fig. 1, it would seem reasonable to associate the observed palaeomagnetic inclination maxima and minima with the passage of the broad negative and positive regions of the non-dipole field beneath the respective sites. Thus one wavelength of the pattern of palaeomagnetic variations would be produced by each revolution of the westward drifting sources located in the outer core. Shorter wavelength variations would be caused by the more rapidly evolving structure within each broad negative and positive region and these should produce variations in the shape of the successive maxima and minima recorded at a given site and also some variation in the shape of any given labelled feature as it is recorded first in Europe and then later in North America.

In principle, we should be able to measure the rate of drift in two ways. First, the time taken for 360° of drift of the whole non-dipole field pattern is given by the time between two successive maxima or minima along the palaeomagnetic records. Figure 4 compares rates of drift through the past 12,000 yr estimated in this way, independently for European data and North American data. Dates of the labelled characteristic features have been taken from the Windermere and St Croix/Kylen records which are the best dated for Europe and North America respectively (Tables 1 and 2). Beyond 9,000 yr BP, dates from the Lac de Joux and Black Sea records have been used for Europe. Features ζ and η , originally identified along the UK records for local correlation purposes are not identified in

this discussion with the major source mechanism and have been ignored for the present analysis.

The above method does not give the sense of drift, but this can be checked as follows. Figures 2 and 3 and Table 1 show that each labelled feature is recorded later in North America than in Europe. This phase lag indicates that the sense of drift was westward and corresponds to the time taken for the centre of a positive or negative region of the non-dipole field to drift under the North Atlantic. There is a 90° longitude difference between the Lake Windermere and St Croix sites. Rates of drift estimated by the two methods are consistent with one another (Fig. 4 and Tables 1 and 2).

It is difficult to estimate quantitative errors for the plotted drift rates. These will depend mainly on the accuracy of the calibration of the palaeomagnetic records, on uniformity of deposition rates between levels dated by radiocarbon and also the accuracy with which the centre of each particular feature may be identified. The overall average drift rate before 2,500 yr BP is estimated at 0.15 deg yr⁻¹ (s.d. = 0.04), 0.16 deg yr⁻¹ (s.d. = 0.02), 0.14 deg yr⁻¹ (s.d. = 0.05) respectively for the European, North American and phase difference data points of Fig. 4.

Note that the characteristic times of the features $\alpha-\epsilon$ are shorter than those of features older than ϵ . The sharp increase inferred for the drift rate since 2,500 yr BP (Fig. 4) results from the assumption that these shorter 'period' features originated from the same mechanism as the older features. The shorter phase differences observed across the Atlantic for features $\alpha-\delta$ (Table 1) is consistent with this assumption. Nevertheless, the possibility that higher degree terms dominated over the quadrupole term between historic times and ~2,500 BP needs to be considered. This is because, although the overall westward sense of drift is confirmed by essentially clockwise looping⁵ of Bauer plots²⁸ of declination against inclination with advancing time for UK since ~7,000 yr BP (ref. 18) and for Lac de Joux since ~11,000 yr BP, a counter-clockwise loop occurs between

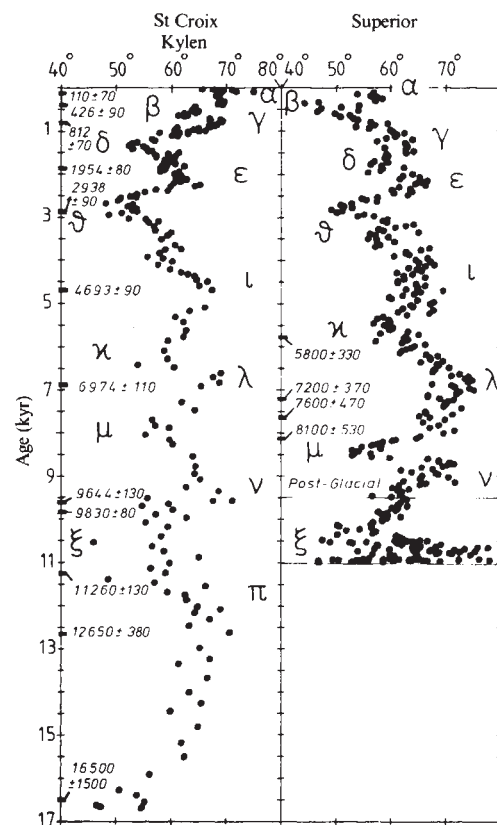


Fig. 3 Palaeomagnetic inclination logs for North American sites. Time scale in uncalibrated radiocarbon years. Radiocarbon ages defining time scale for St Croix and Kylen records²³ and also for the Lake Superior²⁴ record are shown.

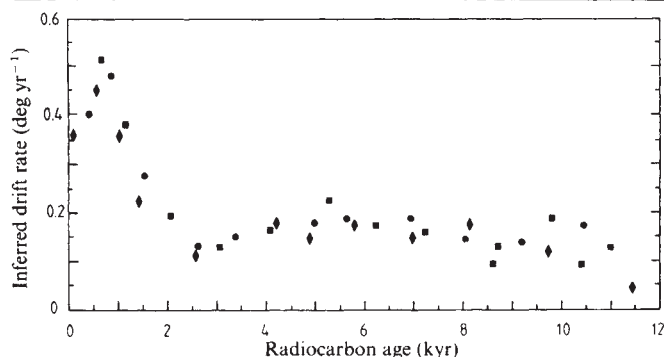


Fig. 4 Inferred westward drift rates through the past 12,000 yr. All ages are uncalibrated radiocarbon ages. Dates of labelled inclination maxima and minima as given by the Windermere and St Croix plots of Figs 2 and 3 have been used back to 9,000 yr BP, thereafter by the Lac de Joux and Kylen plots. ■, European data; ●, North American data; ◆, phase difference.

about 1,500 and 700 yr BP for UK data¹⁸ and also for the Greek data²⁰. According to the Runcorn-Skiles rule⁵ this counter-clockwise looping indicates eastward drifting geomagnetic sources during the time for which we have inferred fast westward drift. However, this rule breaks down under some circumstances²⁹.

Magnitude of the non-dipole field

In the Northern Hemisphere, the broad negative anomaly has gradually increased in strength from -8,000 to -12,000 nT while the positive anomaly has decreased from +12,000 to +8,000 nT through historic time (Fig. 1a). The magnitudes of these anomalies are too small by a factor of about 3 to account

for the observed amplitudes of the broad palaeomagnetic inclination variations. Thus, we infer that the historically observed non-dipole field intensity has been abnormally low relative to the dipole field intensity. Note that the intensity of the dipole has decreased by 5% per century while the quadrupole moment has increased by ~40% per century since AD 1800.

Conclusions

Overall westward drift has persisted throughout the past 12,000 yr. The average rate of 0.14–0.16 deg yr⁻¹ deduced for 12,000–2,500 yr BP is based on a reasonably secure correlation of inclination variations across the North Atlantic. Correlation of the shorter 'period' features α - ϵ is more hazardous and that adopted here must await confirmation from results of a parallel study of declination variations now being carried out. Thus the conclusion that the rate of westward drift increased to about 0.5 deg yr⁻¹ at about 800 yr BP before decreasing to the historical value of 0.25 deg yr⁻¹ should at present be regarded as tentative.

It should be possible soon to answer the question of whether the drifting part of the non-dipole field retained its historical, essentially quadrupolar, symmetry through the post-Glacial time by extending the present studies to Southern Hemisphere sites. Clearly this objective can not be achieved by observations restricted to Northern Hemisphere sites.

A full understanding of long period secular variations requires palaeomagnetic data describing the complete geomagnetic vector as a function of time. As yet, methods developed for the determination of past geomagnetic intensities from the intensity of remanent magnetization of sedimentary rocks have not been properly tested by comparison with archaeomagnetic intensity data from the same locality.

Received 18 March; accepted 6 May 1981.

1. Veinberg, B. P. & Shibaev, V. P. *Transl. No. 0031* (Canadian Department of the Secretary of State, Translation Bureau, 1970).
2. Barraclough, D. R. *Geophys. J. R. astr. Soc.* **36**, 497 (1974).
3. Yukutake, T. *J. Geomagn. Geoelect.* **23**, 11 (1971).
4. Braginskii, S. I. *Geomagn. Aeron.* **12**, 464 and 947 (1972).
5. Skiles, D. D. *J. Geomagn. Geoelect.* **22**, 441 (1970).
6. Bullard, E. C., Freedman, C., Gellman, H. & Nixon, J. *Phil. Trans. R. Soc. A* **243**, 67 (1950).
7. Yukutake, T. & Tachinaka, H. *Bull. Earthq. Res. Inst. Tokyo Univ.* **46**, 1027 (1968).
8. Yukutake, T. & Tachinaka, H. *Bull. Earthq. Res. Inst. Tokyo Univ.* **47**, 65 (1968).
9. James, R. W. *J. Geomagn. Geoelect.* **22**, 241 (1970).
10. Yukutake, T. *PEPI* **20**, 83 (1979).
11. Knapp, D. G. *Geol. Surv. Profess. Pap.* 1118 (1980).
12. Kawai, N. & Hirooka, K. *J. Geomagn. Geoelect.* **17**, 217 (1967).
13. Mackereth, F. J. H. *Limnol. Oceanogr.* **3**, 181 (1958).
14. Molyneux, L. *Geophys. J. R. astr. Soc.* **32**, 479 (1971).

15. Tucker, P. *Geophys. J. R. astr. Soc.* **63**, 149 (1980).
16. Mackereth, F. J. H. *Earth planet. Sci. Lett.* **12**, 332 (1971).
17. Thompson, R. & Turner, G. M. *Geophys. Res. Lett.* **6**, 249 (1979).
18. Turner, G. M. & Thompson, R. *Geophys. J. R. astr. Soc.* **65**, 703 (1981).
19. Creer, K. M., Kroll, E., Readman, P. W. & Tucholka, P. *Geophys. J. R. astr. Soc.* **59**, 287–313 (1979).
20. Creer, K. M., Readman, P. W. & Papamarinopoulos, S. *Geophys. J. R. astr. Soc.* **66**, 193 (1981).
21. Creer, K. M., Hogg, T. E., Readman, P. W. & Reynaud, C. *J. Geophys.* **48**, 139 (1980).
22. Creer, K. M. *Earth planet. Sci. Lett.* **23**, 34 (1974).
23. Banerjee, S. K., Lund, S. P. & Levi, S. *Geology* **7**, 588 (1979).
24. Mothersill, J. S. *Can. J. Earth Sci.* **16**, 1016.
25. Saarnisto, M. *Can. J. Earth Sci.* **12**, 300 (1975).
26. Prest, V. K. *Geol. Surv. Can. Econ. Geol. Rep. No.* 1, 676–764 (1970).
27. Mothersill, J. *Can. J. Earth Sci.* **18**, 448 (1981).
28. Bauer, L. A. *Phys. Rev.* **3**, 34 (1895).
29. Dodson, R. E. *J. geophys. Res.* **84**, 637 (1979).

The *nus* mutations affect transcription termination in *Escherichia coli*

Douglas F. Ward & Max E. Gottesman

Biochemical Genetics Section, National Cancer Institute, Bethesda, Maryland 20205, USA

The nusA1 and nusB5 mutations result in a partial suppression of polarity and thus transcription termination in Escherichia coli. As these mutations block the transcription antitermination activity of bacteriophage λN gene product, they paradoxically seem to enhance transcription termination at phage termination sites. The rho mutation HDF026 displays almost identical properties. These observations suggest that the nusA and nusB gene products may act as termination factors analogous to rho protein.

TERMINATION of transcription serves a dual function in *Escherichia coli*—to keep each transcription unit, or operon, separate and isolated from its neighbours and to allow fine control of expression within the operon structure. The importance of ensuring the isolation of each operon is demonstrated by the interference with expression when transcription enters an operon from without, either in the sense-direction (refs 1, 2 and S. Adhya and M.E.G., in preparation) or in the

opposing direction^{3,4}. The regulation of operon expression by transcription termination within an operon ('attenuation') has been described extensively in the *trp* (refs 5, 6) and *his* (ref. 7) operons.

Premature transcription termination can occur as a result of mutations which introduce nonsense codons into structural genes (see ref. 8) thus exposing transcription termination sites which are not normally active. Such mutations are termed

Table 1 Effect of *nus* mutations on polarity in the *gal* operon

Expt	<i>gal</i>	Galactokinase activity relative to <i>gal</i> ⁺ parent					
		32 °C			42 °C		
		<i>nus</i> ⁺	<i>nusA1</i>	<i>nusB5</i>	<i>nus</i> ⁺	<i>nusA1</i>	<i>nusB5</i>
1	Tam76	0.15	0.20		0.20	0.33	
	TN102	0.06	0.07		0.04	0.05	
	EocB4	0.02	0.02		0.02	0.03	
2	Tam76	0.33		0.50	0.34		0.68
	TN102	0.05		0.09	0.06		0.12

Values represent the ratio of galactokinase (*galK*) activity in strains carrying the indicated *gal* mutation, expressed relative to that in the isogenic *gal*⁺ parent. This normalization eliminates day-to-day variations in the absolute enzyme activities. Activities in *gal*⁺ strains varied from 12 to 29 units at 32 °C, and from 15 to 33 units at 42 °C. Results are the mean of at least two measurements and are accurate to 0.01 (or 10%). As *galK* is distal to both *galT* and *galE*, the level of galactokinase serves as an indication of the polarity of mutations in either of these two genes. Steady-state galactokinase activity was determined in toluenized cell extracts³² and corrected for cell density. Growth medium was L-broth containing 5 mM D-fucose as inducer of the *gal* operon. All strains are derived from *E. coli* K12 strain N4903 (ref. 33). Expt 1 uses a set of isogenic *nusA*⁺ and *nusA1* strains. Expt 2 uses isogenic *nusB*⁺ and *nusB5* strains.

'polar'. DNA insertions into operons usually cause severe polarity; they can introduce transcription as well as translation termination signals.

Termination at many sites requires the action of the *E. coli* rho protein⁹; rho mutants display reduced or no polarity^{10,11}. Mutations in *rpoB* have been isolated^{12,13} which restore polarity in some rho mutant strains. Such allele specificity suggests that interaction between the RNA polymerase β subunit and rho occurs during transcription termination.

Development of bacteriophage λ is regulated at the level of transcription termination. Rho-dependent terminators are present beyond λ genes *N* and *cro*, acting to limit transcription to these immediate-early genes. When *N* product is synthesized, however, these terminators are suppressed, and transcription can extend into the delayed-early gene region. As *N* product suppresses termination at both rho-dependent and rho-independent sites, it may act directly on the transcription termination process, rather than as a simple antagonist of rho¹⁴.

The possibility that pN might interact with RNA polymerase to prevent termination was suggested by the isolation of several mutations in the *rpoB* cistron which blocked pN activity¹⁵⁻¹⁸. However, Greenblatt (personal communication) has failed to demonstrate binding of RNA polymerase to *N* protein *in vitro*; the putative interaction may occur via a nucleic acid or protein intermediary. A phage-specific DNA (or RNA) sequence, *nut*, located before the first terminator, is necessary for pN activity¹⁹. The involvement of additional proteins in the antitermination reaction is suggested by the isolation of other host mutations which prevent *N* function. The *nusA* (ref. 20) and *nusB* (refs 21, 22) mutations do not correspond to any component of RNA polymerase. The *nusA* gene product is apparently identical to protein factor 'L', which is required for optimal expression of β galactosidase in a coupled transcription-translation system²³. *nusA* protein complexes with pN *in vitro*²³; the product of the *nusB* gene has not been identified.

Here we investigate the effect of the *nusA1* and *nusB5* mutations on host and phage physiology. We find that either mutation partially suppresses polarity in *E. coli*, suggesting that the *nusA* and *nusB* gene products may be transcription termination factors.

Effect of *nusA1* and *nusB5* on polarity in the *gal* operon

Suppression of transcription termination by the bacteriophage λ *N* gene product is dependent on several host genes. The host

nusA1 (ref. 20) and *nusB5* (refs 21, 22) mutations are defective for *N* action and restore transcription termination at sites on the phage chromosome. It seemed possible that the *nus* mutations might therefore promote termination in general, in the absence of *N* product, both on viral and bacterial DNA. To test this hypothesis, we measured the polarity of mutations in the *gal* operon in wild-type and *nus* mutant strains. As polarity produced by nonsense mutations, or DNA insertions, results from premature transcription termination, increased polarity would indicate a higher efficiency of termination.

The *galTam76* mutation reduces the expression of *galK* to about 0.15 relative to wild-type levels (Table 1, expt 1). Introduction of the *nusA1* allele elevated the relative galactokinase activity to 0.20. This relief of polarity was slightly more marked at 42 °C; galactokinase levels increased from 0.20 to 0.33 in the *nusA1* strain. Recall that *nusA1* totally blocks *N* activity and hence exerts its maximal termination effect at 42 °C. This result was contrary to expectation; the *nusA1* mutation restores transcription termination in λ and it was therefore expected that *nusA1* would increase polarity in *E. coli*.

Suppression of polarity by *nusA1* was not observed when the strongly polar mutations *galEocB4* and *galTN102*, an IS1 insertion, were assayed (Table 1).

The *nusB5* mutation, like *nusA1*, restricts the growth of λ at high temperature by enhancing termination at sites normally suppressed by the *N* product. As shown in Table 1 (expt 2), the *nusB5* mutation also relieves polarity in the *gal* operon. As was the case with *nusA1*, the limited polarity suppression was observed at both 32 and 42 °C.

The *nusB5* mutation, unlike *nusA1*, reduces the steady-state rate of galactokinase synthesis in *gal*⁺ strains (from 41 to 26 units at 32 °C; from 44 to 25 units at 42 °C). In view of the poor growth rate of strains carrying the *nusB5* mutation (data not shown), the expression of other bacterial operons, including some required for cell growth, may also be adversely affected by *nusB5*.

Interaction of *nus* and *rho* mutations

The experiments presented above suggest that the *nusA* and *nusB* gene products promote transcription termination. To determine whether the *nus* products act independently of rho we investigated whether *nusA* would influence residual polarity in a rho-deficient strain. *galP3* is an IS2 insertion in the *galOP* region²⁴; it introduces a transcription terminator and is highly polar on the *gal* operon⁸. Table 2 shows that in the *rho112 galP3* background, the *nusA1* mutation increased galactokinase activity by more than twofold—from 0.24 to 0.61 at 32 °C, and from 0.30 to 0.85 at 42 °C. Expression of *galK* was similarly stimulated by *nusA1* in a *rho15* mutant.

The *nusA1* mutation does not always enhance polarity suppression by *rho* mutations. Suppression of *galTN102* polarity by *rho112* is not markedly affected by the *nusA1* mutation (Table 2).

Table 2 Suppression of polarity by *rho* and *nusA* mutations

<i>gal</i>	<i>rho</i>	Galactokinase activity relative to <i>gal</i> ⁺ parent			
		32 °C		39 °C	
		<i>nusA</i> ⁺	<i>nusA1</i>	<i>nusA</i> ⁺	<i>nusA1</i>
P3	+	0.02	0.02	0.01	0.02
P3	112	0.24	0.61	0.30	0.85
P3	15	0.51	0.67	0.39	0.75
TN102	+	0.05	0.07	0.02	0.03
TN102	112	0.20	0.22	0.23	0.30
+	112	1.1	1.0	1.2	1.3

Values represent the ratio of galactokinase activity in strains carrying the indicated *gal*, *rho* and *nus* mutations expressed relative to the isogenic *gal*⁺ *rho*⁺ parent. Experimental details are as indicated for Table 1.

Table 3 *trp* attenuation in *nusA1* mutants

	Tryptophan synthetase activity		Attenuation ratio (1)/(2)
	+Attenuator	–Attenuator	
	(1)	(2)	
<i>nusA⁺ rho⁺</i>	11	64	.17
<i>nusA1 rho⁺</i>	14	56	.25
<i>nusA⁺ rho112</i>	10	54	.19
<i>nusA1 rho112</i>	10	51	.20

Tryptophan synthetase³⁴ activities are expressed as units per mg protein, where 1 unit is the amount of enzyme required to convert 0.1 μ mol indole in 20 min at 37 °C. Growth media was M56 minimal medium³⁵ supplemented with glucose (0.2%), casamino acids (0.05%) and thiamine (1 μ g ml⁻¹), plus 100 μ g ml⁻¹ tryptophan. All strains carry the *trpR⁻* mutation. + Attenuator strains are *trpED24* (ref. 36); –attenuator strains are *trpLD102* (ref. 36). The attenuation ratio is accurate to 10%.

Attenuation in *nusA1* mutant strains

The apparent site-specificity of polarity suppression by *nusA1* prompted us to examine transcription termination at the attenuator of the *trp* operon. Several *rho* mutations have been found to increase the expression of *trp* enzymes²⁵, suggesting a possible role of *rho* in *trp* attenuation.

Two sets of strains carrying combinations of *nusA1* and *rho112* were constructed. The first carried the *trpLD102* deletion which removes the attenuator; the second carried *trpED24*, a deletion similar in size and extent to *trpLD102* but which leaves the attenuator intact. *trp* expression was genetically derepressed by the introduction of a *trpR* mutation.

The effects of *rho* and *nusA1* mutations on attenuation were determined by assaying tryptophan synthetase, the product of the *trpA* and *trpB* genes (Table 3). Unlike other *rho* mutant alleles tested, *rho112* does not significantly reduce *trp* operon attenuation. By itself, the *nusA1* mutation does cause a significant, although small, suppression of *trp* attenuation. This effect was not apparent in the *rho112* background.

Interactions of *nusA* and *rho* mutations affect the growth of *E. coli* and λ

The *rho15* and *rho112* mutants both fail to grow in rich media at temperatures above 39 °C. In view of the pleiotropic properties of these *rho* mutations, it has not been possible to ascribe this conditional lethality to failure of transcription termination and, in fact, thermoresistant survivors of *rho15* mutants in which polarity has not been restored have been isolated²⁶.

It was surprising to observe that a *rho112 nusA1* double mutant grew well at 40 °C (Table 4). In contrast to a singly mutant strain, *nusA1* or *rho112*, the *nusA1 rho112* double mutant fails to form colonies at 32 °C. Unlike *nusA1*, the *nusB5* mutation neither suppresses the temperature sensitivity of *rho112* mutants nor renders them cryosensitive.

With respect to growth of λ the *rho112* mutation is quite clearly epistatic to *nusA1* or *nusB5*. At 37 °C the growth of λ C₁₇ is blocked by *nusA1* or *nusB5*; phage growth is restored by the introduction of *rho112* into the *nus* mutant strains (Table 4). In addition, the double mutant supports the growth of λ N⁻. Thus, the ability of the *nus* mutations to make a phage termination site resistant to *N* function is dependent on an active *rho* allele.

Discussion

We have presented evidence which implies that the *nus* mutations of *E. coli* have opposite effects in phage λ and bacteria. In λ , the products of the *nus* genes are required for the phage *N* function to act and therefore λ fails to grow in *nus* mutant strains. As the role of *N* function is to prevent termination, the *nus* mutations can be considered to enhance transcription termination on λ DNA. However, in *E. coli*, the polarity of certain mutations is partially suppressed by the *nus* mutations.

As polarity is caused by premature termination of transcription, the *nus* mutations seem to reduce termination in *E. coli*.

The degree of polarity suppression by *nusA1* and *nusB5* depends on the polar mutation tested. Partial suppression by *nusA1* occurs in the case of weakly polar *gal* mutations, but little or no relief of polarity is seen with more polar mutations in the *gal* operon such as *galEocB4*, IS1 or IS2 insertions. *nusB5* is more effective than *nusA1* in suppressing polarity and shows some relief of polarity caused by the IS1 insertion mutant, *galTN102*, particularly at 42 °C. The most marked effect of the *nusA1* mutation on polarity is in conjunction with *rho112*, a *rho* mutation which has significant residual *rho* activity. Here, the *nusA1* and *rho112* mutations act synergistically to suppress a highly polar IS2 insertion.

The *nusA1* mutation produces a partial relief of attenuation in the *trp* operon. An apparent stimulation by *nusA* protein of the *in vitro* synthesis of RNA polymerase β and β' subunits has been interpreted as suggesting that *nusA* protein suppresses an attenuator that regulates $\beta\beta'$ expression^{23,27,28}. We have not, however, observed any significant effect of the *nusA1* mutation on the synthesis of these RNA polymerase subunits *in vivo* as judged by the incorporation of ¹⁴C-labelled amino acids and visualization of the proteins in SDS-polyacrylamide gels (data not presented).

We have also tested for other interactions between the *nus* mutations and *rho*. With respect to phage growth, *rho* mutations are clearly epistatic to *nusA1* or *nusB5*; λ C₁₇, which fails to form plaques on *nus⁻* hosts, grows well on a *rho⁻ nusA1* or a *rho⁻ nusB5* double mutant. For bacterial growth, *nusA1* is epistatic to *rho⁻*; a *rho⁻* mutant which fails to form a colony at 40 °C will do so when a *nusA1* mutation is introduced into the strain. The double mutant is, however, cryosensitive and does not survive at 32 °C. Introduction of the *nusB5* allele into a *rho⁻* mutant does not alter its growth pattern; it remains thermosensitive and cryoresistant.

The thermosensitivity of *rho⁻* strains is not well understood. The *rho15* and *rho112* mutant proteins are thermolabile *in vitro*. Yet the isolation of secondary mutations which suppress the thermosensitivity of the *rho* mutants, without restoring polarity (at least to those sites tested), suggests that failure of transcription termination may not be the lethal event (ref. 29 and A. Das, personal communication). Furthermore, the wide range of defects in the *rho⁻* mutants—abnormalities in DNA synthesis, in recombination and in oxidative phosphorylation—presents several potential causes for thermosensitivity. The observation that the *nusA1* mutation suppresses the thermosensitivity of *rho⁻* mutants in addition to enhancing suppression of certain polar mutations is not inconsistent with the above analysis. We can imagine that *nusA1* affects the pleiotropic character which is responsible for cell death at 40 °C by a mechanism unrelated to transcription termination. Alternatively, the *nusA* gene protein may act at some bacterial terminators, vital for cell growth, as it does at λ terminators, by stimulating transcription termination.

The *rho* mutant, *rhoHDF026*, has been characterized and seems to be phenotypically very similar to *nusA1* and *nusB5*—polarity suppression is partial and the activity of the λ N function is reduced³⁰). The similarity between these mutants is consistent with the idea that *nusA* and *nusB*, like *rho*, are transcription termination factors.

Table 4 Interactions of *rho*, *nusA1* and *nusB5* mutations

<i>rho</i>	<i>nus</i>	Cell growth		Phage plating		
		32 °C	42 °C	λ in ₅	λ C ₁₇	λ N ⁻
+	+	+	+	+	+	–
+	A1	+	+	+	–	–
+	B5	+	+	+	–	–
112	+	+	–	+	+	+
112	A1	–	+	+	+	+
112	B5	+	–	+	+	+

In an accompanying article³¹, Greenblatt *et al.* present evidence that *nusA* protein acts *in vitro* as a site-specific transcription termination factor. His results are consistent with our *in vivo* experiments which indicate that the *nusA1* mutation partially relieves polarity in the *gal* operon. The incomplete relief of polarity by *nusA1* is best explained by assuming that *nusA1* protein retains considerable termination activity. The paucity of *nusA* mutants suggests that *nusA* may have a vital function. λ N protein interacts with *nusA* protein *in vitro*²³, and presumably *in vivo*, to antagonize *nusA* termination function. It is possible that the *nusA1* mutation primarily affects the ability of *nusA* protein to interact with *N* rather than affecting its transcription termination activity.

The inhibition of λ growth by *nusA1* is overcome by introducing a mutation in the *rho* gene, indicating that transcription

termination caused by *nusA1* must be occurring at rho-dependent sites. We cannot eliminate the possibility that other terminators may exist which are dependent solely on *nusA* protein.

We will be presenting elsewhere evidence that the *nusA* and *nusB* gene products interact. It is not unreasonable, therefore, that the *nusA1* and *nusB5* mutants should have a similar phenotype.

Further investigation of the role of *nusA* and *nusB* proteins in bacterial and phage metabolism requires the isolation of additional *nus* mutations, of secondary mutations which suppress the *Nus*⁻ phenotype and of reproduction of *nus* protein action *in vitro*. This work is in progress.

We thank Don Court and Sankar Adhya for many helpful discussions.

Received 5 December 1980; accepted 14 April 1981.

- Merrill, C., Gottesman, M. E., Court, D. L. & Adhya, S. *J. molec. Biol.* **118**, 241–245 (1978).
- Hausler, B. & Somerville, R. L. *J. molec. Biol.* **127**, 353–356 (1979).
- Levinthal, M. & Nikaido, H. *J. molec. Biol.* **42**, 511–520 (1969).
- Ward, D. F. & Murray, N. E. *J. molec. Biol.* **133**, 249–266 (1979).
- Bertrand, K. *et al. Science* **189**, 22–26 (1975).
- Keller, E. B. & Calvo, J. M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 6186–6190 (1979).
- Johnston, H. M., Barnes, W. M., Chumley, F. G., Bossi, L. & Roth, J. R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 508–512 (1980).
- Adhya, S. & Gottesman, M. E. *Rev. Biochem.* **47**, 967–996 (1978).
- Roberts, J. *Nature* **224**, 1168–1174 (1969).
- Richardson, J. P., Grimley, C. & Lowery, C. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1725–1728 (1975).
- Das, A., Court, D. & Adhya, S. *Proc. natn. Acad. Sci. U.S.A.* **73**, 1959–1963 (1976).
- Das, A., Merrill, C. & Adhya, S. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4828–4832 (1978).
- Guarente, L., Mitchell, D. & Beckwith, J. *J. molec. Biol.* **112**, 423–436 (1977).
- Adhya, S., Gottesman, M. & deCrombrughe, B. *Proc. natn. Acad. Sci. U.S.A.* **71**, 2534–2538 (1974).
- Baumann, M. F. & Friedman, D. I. *Virology* **73**, 128–138 (1976).
- Georgopoulos, C. P. *Proc. natn. Acad. Sci. U.S.A.* **68**, 2977–2981 (1971).
- Ghyben, A. & Pironio, M. *J. molec. Biol.* **65**, 259–272 (1972).
- Sternberg, N. *Virology* **73**, 139–154 (1976).
- Salstrom, J. S. & Szybalski, W. *J. molec. Biol.* **124**, 195–221 (1978).
- Friedman, D. I. *The Bacteriophage Lambda* (ed. Hershey, A. D.) 733–738 (Cold Spring Harbor, New York, 1971).
- Friedman, D. I., Baumann, M. & Baron, L. S. *Virology* **73**, 119–127 (1976).
- Georgopoulos, C. P. *et al. Molec. gen. Genet.* **179**, 55–61 (1980).
- Greenblatt, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 1991–1994 (1980).
- Adhya, S. & Shapiro, J. A. *Genetics* **62**, 231–247 (1969).
- Korn, L. & Yanofsky, C. *J. molec. Biol.* **103**, 395–409 (1976).
- Das, A., Court, D. & Adhya, S. *Molecular Basis of Host Virus Interactions* (ed. Chakravarty, M.) 459–467 (Science Press, Marikville, 1978).
- Zaruchi-Schulz, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 6115–6119 (1979).
- Barry, G., Squires, C. L. & Squires, C. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4922–4926 (1979).
- Guidi-Rontani, C., Danchin, A. & Ullmann, A. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5799–5801 (1980).
- Simon, L. D., Gottesman, M., Tomczak, K. & Gottesman, S. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1623–1627 (1979).
- Greenblatt, J., McLimont, M. & Hanly, S. *Nature* **292**, 215–220 (1981).
- Sherman, J. R. & Adler, J. *J. biol. Chem.* **238**, 873–878 (1963).
- Gottesman, M. E., Adhya, S. & Das, A. *J. molec. Biol.* **140**, 57–75 (1980).
- Smith, O. H. & Yanofsky, C. *Meth. Enzym.* **5**, 801–802 (1962).
- Gottesman, M. E. & Yarmolinsky, M. B. *J. molec. Biol.* **31**, 487–505 (1968).
- Jackson, E. & Yanofsky, C. *J. molec. Biol.* **71**, 149–161 (1972).

Termination of transcription by *nusA* gene protein of *Escherichia coli*

Jack Greenblatt, Marjorie McLimont & Sarah Hanly

Banting and Best Department of Medical Research, University of Toronto, Toronto, Canada M5G 1L6

The nusA gene protein of Escherichia coli and N gene protein of bacteriophage λ interact in vitro and cooperate in vivo to prevent transcription termination. In vitro the nusA gene protein causes RNA polymerase to pause in the t_{R2} terminator region of λ DNA. A completed termination event at t_{R2} requires both the nusA gene protein and the previously described E. coli termination factor rho. The nusA gene protein is therefore both a transcription termination factor and a protein which couples antitermination factors to the elongating transcription complex.

BEYOND the ends of the first genes *N* and *cro* of the P_L and P_R early operons of bacteriophage λ are the transcription termination sites t_{L1} and t_{R1} (see Fig. 1). Termination of transcription occurs at these sites *in vitro* or *in vivo* only in the presence of the transcription termination factor rho^{1–5}. In an infection of rapidly growing *Escherichia coli* bacteria by phage λ the *E. coli* RNA polymerase is able to transcribe λ genes distal to the *N* and *cro* genes because the product of the *N* gene of the phage prevents termination of transcription at t_{L1} and t_{R1} (refs 2–11). The *N* gene protein also prevents termination of transcription at other more distal termination sites in both of the early operons^{4,12–14}. It is thought that *N* protein recognizes particular sites, called *nut* sites, in the nucleic acid of the λ early operons^{5,15} and somehow modifies RNA polymerase molecules at these sites so that they will resist termination of transcription at most termination sites^{16–20}.

Friedman and Georgopoulos and their co-workers (and others) have isolated *E. coli* mutants in which phage λ cannot grow because its *N* gene protein cannot prevent transcription

termination^{21–24}. One such mutation is the *nusA1* mutation at 68 min on the standard *E. coli* map, which blocks λ growth at 42 °C, but not at 30 °C (refs 21, 25). The *nusA1* mutant protein does not inactivate *N* protein because the *nusA*⁺ allele is dominant to the *nusA1* allele²⁵. This suggests that the *nusA* protein is involved in preventing termination of the transcription of λ DNA. That a phage λ *N* gene mutant, λ NpnA, grows on *nusA1* bacteria at 42 °C (ref. 26) suggests that the *nusA* gene protein might interact directly with the *N* gene protein. Recently, the *nusA* gene protein has been identified as a 69,000-dalton acidic protein which binds directly to *N* protein²⁷.

The *nusA* gene protein is identical to L factor²⁸, a protein which was purified as an activity necessary for DNA-dependent β -galactosidase synthesis *in vitro*²⁹. L factor also stimulates the DNA-dependent synthesis *in vitro* of the β and β' subunits of *E. coli* RNA polymerase³⁰. These observations can perhaps be rationalized by noting that there is a rho-sensitive transcription termination site in the middle of the *lacZ* gene³¹ and a probable transcriptional attenuator proximal to the *rpoB* gene³². These

results suggest that the *nusA* gene protein may have a role in overcoming transcription termination at bacterial terminators.

We have now examined the effects of wild-type and mutant *nusA* gene proteins on the transcription *in vitro* of λ DNA by *E. coli* RNA polymerase. Our results show that the *nusA* gene protein can act to terminate the transcription of the P_R operon of λ . Other experiments, to be published elsewhere³³, demonstrate that the *nusA* protein binds specifically, directly and with equimolar stoichiometry to the transcribing form (core enzyme) of RNA polymerase. The *nusA* gene protein is therefore both a termination factor for RNA synthesis and a protein which can couple transcription antitermination factors, such as the λ *N* protein, to the transcription complex.

Inhibition of transcription *in vitro* by the *nusA* gene protein

In our first preparations of the *nusA* gene protein we followed the L factor purification scheme of Kung *et al.*²⁹ and pooled those fractions containing a protein with the same SDS-polyacrylamide gel mobility and isoelectric point as a radiolabelled *nusA* gene protein standard²⁷. We then noted that preparations of the *nusA* gene protein at least 90% pure contain an activity that inhibits the transcription *in vitro* of λ DNA by pure *E. coli* RNA polymerase (Fig. 2). The dose-response curve is similar to that produced by rho factor¹ in that the inhibition never becomes greater than about 50% even at very high concentrations of *nusA* protein (Fig. 2). Moreover, as the *nusA* gene protein does not inhibit transcription in reactions carried out in the presence of rho factor (Fig. 2), our results suggested that the *nusA* gene protein might be acting to terminate the transcription of λ DNA at a site or sites located either at or distal to the rho-sensitive terminators.

A consideration of the genetic evidence dealing with the *nusA* gene²¹ and of the biochemical data on L factor^{29,30} suggested that the *nusA* gene protein was involved in preventing the termination of transcription (see above). The idea that it might be causing the termination of transcription was therefore sufficiently unexpected for us to seek convincing evidence that the apparent termination activity present in our purified *nusA* protein was indeed a property of the *nusA* protein itself and not of a contaminant present in the preparation.

Purification of wild-type and mutant *nusA* gene proteins

Figure 3a,c,e demonstrates that an activity which inhibits λ transcription *in vitro* has the same chromatographic behaviour as the *nusA* protein on DEAE-cellulose (Fig. 3a) and hydroxylapatite (Fig. 3c) and the same sedimentation velocity as the *nusA* protein during centrifugation in a glycerol density gradient (Fig. 3e). In the purification scheme illustrated in Fig. 3 we modified the L factor purification procedure of Kung *et al.*²⁹ by using a buffer of higher ionic strength during the high speed centrifugation step and by replacing their gel filtration column

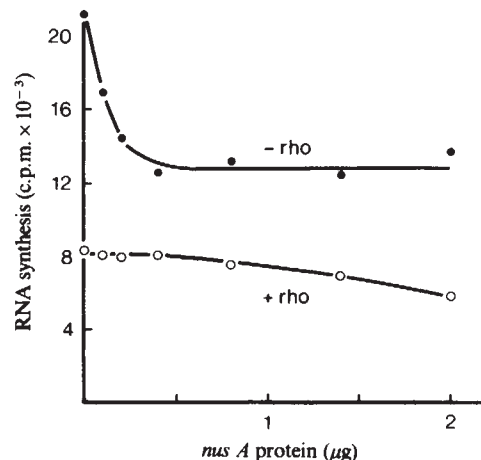


Fig. 2 Effect of the *nusA* protein on total λ transcription in the absence and in the presence of rho factor. Each reaction contained, in a total volume of 50 μ l, 20 mM Tris-HCl, pH 7.9, 80 mM KCl, 5 mM MgCl₂, 1 mM dithiothreitol, 100 μ M EDTA, 200 μ M each of ATP, CTP and GTP, 50 μ M ³H-UTP at 16,000 c.p.m. per nmol 2 μ g λ C1857 DNA, 1 μ g *E. coli* RNA polymerase holoenzyme⁴⁸, the indicated amount of *nusA* protein, and either no rho factor (●) or 50 ng rho factor (○). Incubation was for 30 min at 37 °C and ³H-UTP incorporated into RNA was determined using DE81 filters⁴⁹.

with a hydroxylapatite column. In our hands, at least, the result is a higher yield of *nusA* protein with a greater degree of purity. The fact that an activity that inhibits transcription of λ DNA *in vitro* co-purifies with the *nusA* protein is consistent with the idea that such an activity is indeed a property of the *nusA* protein.

To establish this point more securely we purified the *nusA* protein from two strains, *nusA1* (ref. 21) and *groN785* (ref. 22), that each has a *nusA* protein with a modified isoelectric point²⁷. We were able to do so because the *nusA* mutation in *groN785* seems to be a silent mutation²⁷ and because the *nusA1* mutation is a conditional defect that is very weak at low temperature²¹. Presumably as a result of the charge change or a configurational change introduced by the mutation, each of the mutant *nusA* proteins behaves somewhat differently from the wild-type protein during chromatography on DEAE-cellulose and hydroxylapatite. The entire purification of the *nusA* protein from the *groN785* strain is illustrated in Fig. 3b,d. Both mutant proteins elute later than the wild-type protein from columns of both DEAE-cellulose and hydroxylapatite.

Our purification scheme leads to a high degree of purity (>90%) for the wild-type *nusA* protein (Fig. 3e) and for both mutant *nusA* proteins (see Fig. 3d for *nusA785*). All three proteins preserve their isoelectric homogeneity throughout the purification (not shown). Most importantly, both the mutant *nusA785* protein (Fig. 3b,d) and the mutant *nusA1* protein (not shown) co-purify with an activity that inhibits λ transcription *in vitro*, just as does the wild-type protein (Fig. 3a,c,e). As the behaviours during purification of the three allelic proteins are non-identical, the ability to inhibit transcription of λ DNA *in vitro* is therefore a genuine property of the *nusA* gene protein.

Termination of λ transcription *in vitro* at t_{R2} by the *nusA* gene protein

We have used gel electrophoresis to examine the effect of the *nusA* gene protein on λ transcripts made *in vitro* in the presence of [α -³²P]UTP. We chose an electrophoresis system used by Roberts³⁴ to analyse λ transcripts made in reactions containing RNA polymerase and no additional factors. Roberts noted that most RNA polymerase molecules transcribing the P_R operon synthesize a large 2.3×10^6 -dalton transcript (see Figs 1, 4a). His mapping data suggested that this long transcript was probably co-terminal, at t'_R , with the short 6×10^4 dalton RNA transcribed *in vitro* from the late promoter P'_R (Figs 1, 4a).

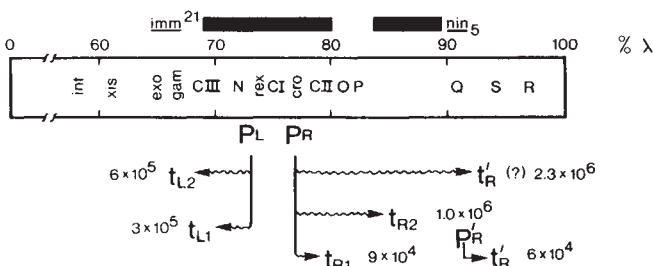


Fig. 1 Partial genetic and transcriptional map of bacteriophage λ . The sizes of the P_R - t_{R1} and P'_R - t'_R transcripts have been determined by sequencing⁴⁵⁻⁴⁷ and the indicated sizes, in daltons, of the other transcripts estimated by electrophoresis^{5,34}. For simplicity, the known λ transcripts beginning at pr_m , pre , P_O and P_I have been omitted from the diagram. *nin5* is a deletion and *imm21* is a deletion-substitution.

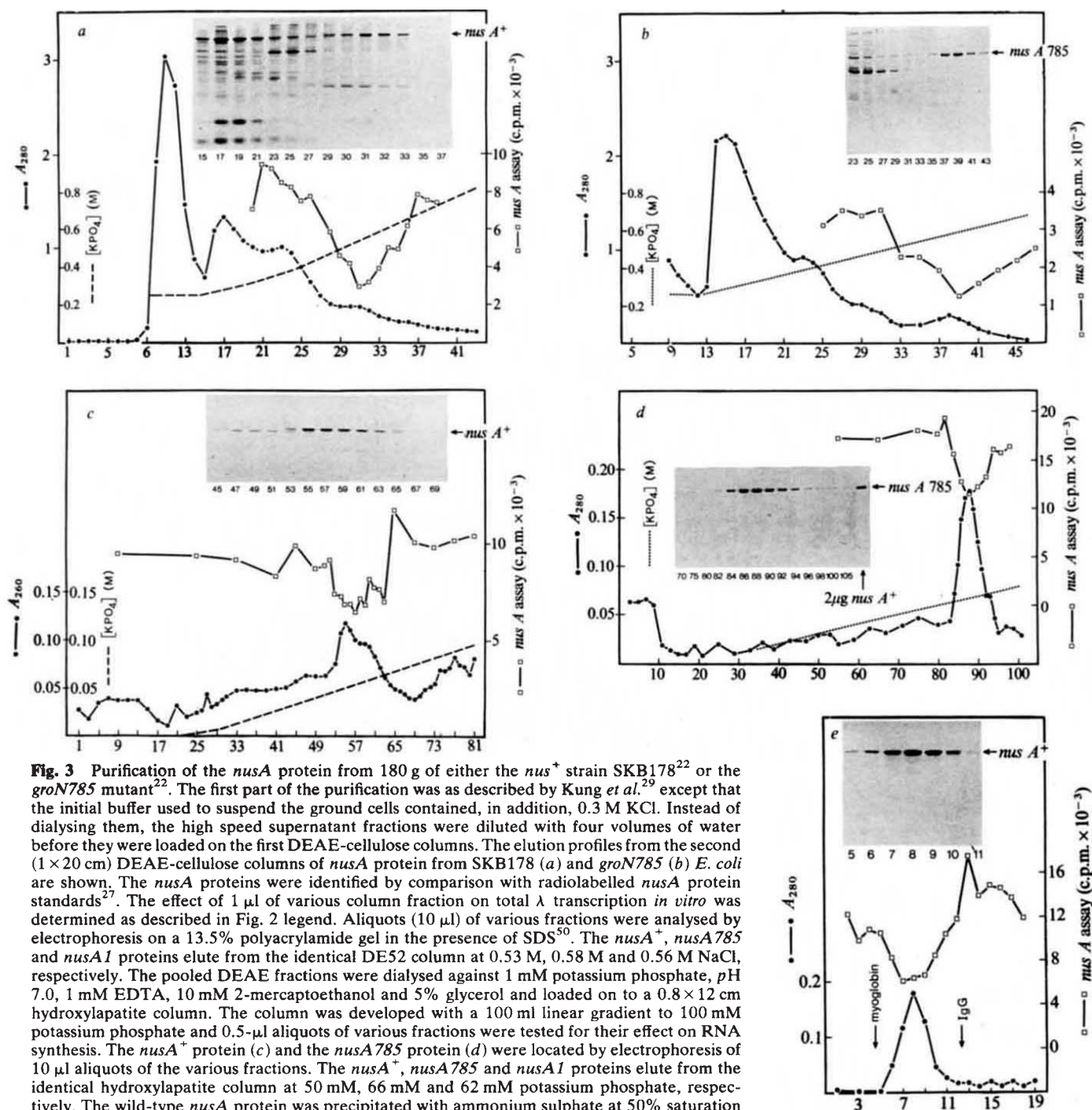


Fig. 3 Purification of the *nusA* protein from 180 g of either the *nus*⁺ strain SKB178²² or the *groN785* mutant²². The first part of the purification was as described by Kung *et al.*²⁹ except that the initial buffer used to suspend the ground cells contained, in addition, 0.3 M KCl. Instead of dialysing them, the high speed supernatant fractions were diluted with four volumes of water before they were loaded on the first DEAE-cellulose columns. The elution profiles from the second (1 × 20 cm) DEAE-cellulose columns of *nusA* protein from SKB178 (a) and *groN785* (b) *E. coli* are shown. The *nusA* proteins were identified by comparison with radiolabelled *nusA* protein standards²⁷. The effect of 1 μ l of various column fraction on total λ transcription *in vitro* was determined as described in Fig. 2 legend. Aliquots (10 μ l) of various fractions were analysed by electrophoresis on a 13.5% polyacrylamide gel in the presence of SDS⁵⁰. The *nusA*⁺, *nusA785* and *nusA1* proteins elute from the identical DE52 column at 0.53 M, 0.58 M and 0.56 M NaCl, respectively. The pooled DEAE fractions were dialysed against 1 mM potassium phosphate, pH 7.0, 1 mM EDTA, 10 mM 2-mercaptoethanol and 5% glycerol and loaded on to a 0.8 × 12 cm hydroxylapatite column. The column was developed with a 100 ml linear gradient to 100 mM potassium phosphate and 0.5- μ l aliquots of various fractions were tested for their effect on RNA synthesis. The *nusA*⁺ protein (c) and the *nusA785* protein (d) were located by electrophoresis of 10 μ l aliquots of the various fractions. The *nusA*⁺, *nusA785* and *nusA1* proteins elute from the identical hydroxylapatite column at 50 mM, 66 mM and 62 mM potassium phosphate, respectively. The wild-type *nusA* protein was precipitated with ammonium sulphate at 50% saturation and then further purified by centrifugation for 42 h at 40,000 r.p.m. in a Beckman SW 40 rotor through a 10–30% (W/V) glycerol gradient containing 10 mM potassium phosphate, pH 7.0, 100 mM NaCl, 1 mM EDTA and 10 mM 2-mercaptoethanol (e). 1- μ l Aliquots of gradient fractions were tested in transcription reactions and 5- μ l aliquots were analysed by electrophoresis. The *nusA* gene protein sediments at 4.2S. The yields of the wild-type and *groN785* *nusA* proteins were, respectively, about 2 mg and 4 mg per 200 g wet weight of *E. coli*. Each was more than 90% pure, as was the *nusA1* protein, purified in a similar manner (not shown).

Some of the RNA polymerase molecules transcribing the *P_R* operon seem to stop at a site *t_{R2}* located about midway between *t_{R'}* and the rho-dependent terminator *t_{R1}* (ref. 34 and Fig. 1). The size of the resulting transcript is 1.0×10^6 daltons (Fig. 4a). The *t_{R2}*-terminated transcript is easily identified as the one which is eliminated by the *nin*₅ deletion (Fig. 4h), which deletes at least one *P_R* operon terminator *in vivo*^{12,35}. The *P_R* operon transcript made from the λ *nin*₅ template is the *P_R*-*t_{R'}* transcript shortened to 1.6×10^6 daltons by the length of the *nin*₅ deletion. Experiments in which rifampicin was used to synchronize initiation of transcription and then gel electrophoresis to monitor the kinetics of RNA chain elongation (not shown) have indicated that the RNA polymerase molecules pause for an average of

1–2 min at *t_{R2}*; ultimately, all of them synthesize a large 2.3×10^6 -dalton RNA.

The effect of the *nusA* gene protein on λ transcription *in vitro* is shown in Fig. 4b. It eliminates the long *P_R*-*t_{R'}* transcript and leaves the shorter *P_R*-*t_{R2}* transcript. As shown in Fig. 4i, it does not eliminate the *P_R*-*t_{R'}* transcript when λ *nin*₅, a template which lacks *t_{R2}*, is transcribed. Thus, the major effect of the *nusA* gene protein that can be observed by gel electrophoresis is to enhance the efficiency to apparent transcription termination at *t_{R2}*.

The molar yield of the *P_R*-*t_{R2}* transcript made in the presence of the *nusA* protein is somewhat less than expected based entirely on a termination model (that is, [*P_R*-*t_{R2}*] in the presence of *nusA* < [*P_R*-*t_{R2}*] + [*P_R*-*t_{R'}*] in the absence of *nusA*). This is

because RNA polymerase molecules arrive nonsynchronously at t_{R2} , because the *nusA* protein increases the transit time for RNA polymerase from P_R to t_{R2} by about 50%, and because the *nusA* protein decreases the rate of RNA chain initiation by about 20% (J. G., unpublished data). These other effects of the *nusA* protein on λ transcription *in vitro* can be accounted for by the tight association of the *nusA* protein with core RNA polymerase at sites other than terminator sites³³ and by competition between the *nusA* protein and the RNA polymerase σ subunit for binding to core RNA polymerase³³. The experiment of Fig. 2 suggests that the effect of the *nusA* protein on transcription initiation is, in any case, only minor.

Figure 4e demonstrates that the *nusA* gene protein is a termination factor that, to be effective, must be present during the transcription reaction. When transcription is carried out in the absence of the *nusA* gene protein, subsequent incubation in the presence of the *nusA* gene protein and the transcriptional inhibitor streptolydigin does not result in the conversion of the P_R - t_R' transcript to the shorter P_R - t_{R2} transcript. Therefore, the *nusA* gene protein is not a ribonuclease that processes the P_R - t_R' transcript to produce a shortened derivative. Kinetic experiments (not shown) have indicated that the *nusA* protein does not cause irreversible chain termination at t_{R2} , but instead causes RNA polymerase to pause for 10–15 min at several closely linked sites in the t_{R2} region. Irreversible termination at t_{R2} requires both the *nusA* protein and a low concentration of rho factor ($3 \mu\text{g ml}^{-1}$, a concentration which does not cause termination at t_{R2} in the absence of the *nusA* protein).

In the absence of the *nusA* protein and rho factor, RNA polymerase synthesizes a 6×10^5 dalton transcript from the P_L operon, terminating at t_{L2} (ref. 34 and Figs 1, 4a). Kinetic experiments have indicated that the event at t_{L2} is a pause of very long, but not indefinite, duration (about 10 min) which is not affected by the *nusA* protein. Hybridization of λ RNA synthesized *in vitro* to various fragments of λ DNA (J. G., unpublished data) confirmed that the *nusA* protein inhibits the synthesis of distal RNA from the P_R operon, but has little effect on the synthesis of distal RNA from the P_L operon (unlike rho, which inhibits both). This suggests that, at least *in vitro*, the *nusA* protein acts as a strong termination factor for the P_R operon, but not for the P_L operon.

Other large λ RNAs are made in the absence of the *nusA* gene protein and not in its presence (see Fig. 4), and are particularly evident when the *nin5* deletion moves the long t_R' -terminated transcript lower in the gel (Fig. 4h, i). Experiments with λ imm²¹ DNA (Fig. 4f, g) suggest that some of them come from the immunity region, but they have not been studied in any greater detail.

Termination of transcription *in vitro* at t_{R2} by mutant *nusA* gene proteins

To show that termination at t_{R2} is caused by the *nusA* gene protein and not by a contaminant present in our preparation, we carried out similar experiments with *nusA* protein derived from the *groN785* and *nusA1* strains^{21,22,27}. Thus, *nusA785* protein (Fig. 4l) is just as effective as the *nusA*⁺ protein (Fig. 4k) in terminating transcription at t_{R2} , as is consistent with the lack of phenotype conferred on λ or *E. coli* growth by the *nusA785* mutation²⁷. In fact, the *nusA785* protein is just as effective as the *nusA*⁺ protein at t_{R2} even when the mutant RNA polymerase of the *groN785* strain^{22,36} is used in the experiment (Fig. 4m–o). As the *nusA785* protein behaves differently during purification from the *nusA*⁺ protein (it elutes later from both DEAE cellulose and hydroxylapatite; see Fig. 3), it must be the *nusA* gene protein itself that causes termination at t_{R2} .

The mutant *nusA1* protein does behave differently from the wild-type protein (Fig. 5). It is able to cause termination at t_{R2} , but does so with a specific activity several-fold lower than normal. This is true even in reactions carried out at 30 °C and the effect is only slightly accentuated in reactions carried out at higher temperatures. The lack of a strong temperature effect on *nusA1* termination activity *in vitro* is not unexpected: *in vivo*

experiments have suggested that the *nusA1* mutant protein is only temperature sensitive for antitermination²¹, not for termination. The low specific activity of the *nusA1* protein is not a consequence of its having been almost totally denatured during purification because the pure *nusA1* protein is fully capable of binding to λ N gene protein and to RNA polymerase (J. G. and Joyce Li, unpublished data). It seems likely that the reduced activity of the *nusA1* mutant protein is adequate or almost adequate *in vivo* for terminating transcription (see accompanying article³⁷), but inadequate for coupling the N gene antitermination protein to the transcription complex. The fact that the *nusA1* protein has an altered t_{R2} termination activity provides additional evidence that it is really the *nusA* gene protein that causes termination at t_{R2} .

Termination of transcription *in vitro* at t_{R2} by a mutant RNA polymerase

The experiment shown in Fig. 4p–s suggests that the termination effect of the *nusA* protein *in vitro* represents a real physiological phenomenon, and is not simply an artefact of our conditions for transcription *in vitro*. We purified RNA polymerase from an *E. coli* strain isolated by Lecocq and Dambly³⁸ which has a mutation, *rif501*, conferring resistance to rifampicin in the β subunit of its RNA polymerase. One of the properties of this strain is its ability to support λ growth at high temperature in the absence of a functional N gene protein³⁸. The experiments of Lecocq and Dambly³⁸ suggest that the *rif501* RNA polymerase is resistant *in*

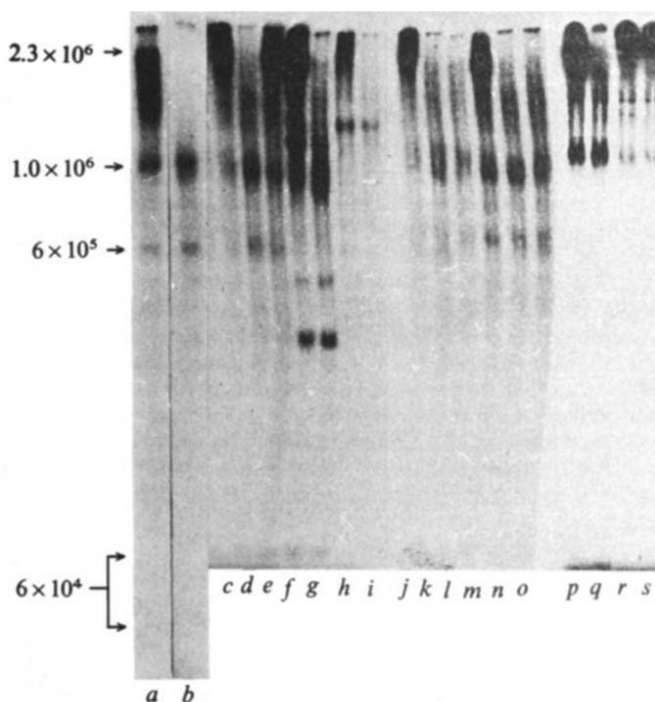


Fig. 4 Electrophoretic analysis of λ transcripts made *in vitro*. Transcription reactions were carried out as described in Fig. 2 legend except that they contained [α -³²P]UTP at 200 μCi per μmol instead of ³H-UTP. Reactions contained the indicated DNAs and, where indicated, 0.5 μg *nusA* protein. Reactions were stopped with 100 μl of 9 M urea, 1% SDS and aliquots were electrophoresed on a slab gel containing 0.5% agarose, 2.5% polyacrylamide and 0.1% SDS³⁴. a, λ DNA; b, λ DNA, *nusA*⁺ protein; c, λ DNA; d, λ DNA, *nusA*⁺ protein; e, λ DNA, reaction stopped with 10 $\mu\text{g ml}^{-1}$ streptolydigin and incubated for another 30 min in the presence of *nusA*⁺ protein; f, λ imm²¹ DNA; g, λ imm²¹ DNA, *nusA*⁺ protein; h, λ nin⁵ DNA; i, λ nin⁵ DNA, *nusA*⁺ protein; j, λ DNA; k, λ DNA, *nusA*⁺ protein; l, λ DNA, *nusA785* protein; m, n, o, same as j, k, l except RNA polymerase from the *groN785* strain³⁶ instead of wild-type RNA polymerase; p, λ DNA, q, λ DNA, *nusA*⁺ protein; r, s, same as p, q except that the core component⁵¹ of the RNA polymerase was from the *rif501* mutant³⁸ instead of from wild-type *E. coli*.

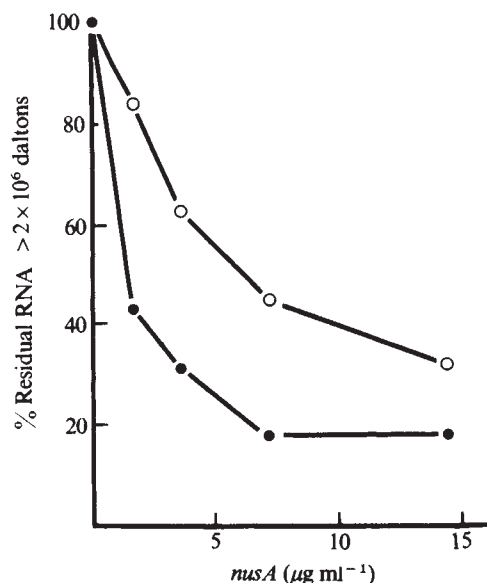


Fig. 5 Comparison of the effects of wild-type *nusA* protein and the *nusA1* mutant protein on λ transcription *in vitro*. Reactions containing the indicated concentrations of the *nusA*⁺ protein (●) or the *nusA1* protein (○) were carried out and the products analysed as in Fig. 4. The high molecular weight regions of the gel tracks were cut out and counted and the results expressed as a percentage of the counts obtained in tracks of reactions containing no added *nusA* protein.

in vivo to termination at terminators of the P_R operon of λ , but terminates normally at the terminators of the P_L operon of λ . As the *nusA* gene protein is active *in vitro* on the transcription of the P_R operon, but not on the P_L operon, it seemed likely that the *rif501* RNA polymerase was resistant to the *nusA* protein. As shown in Fig. 4r, s the *rif501* RNA polymerase does not respond *in vitro* at t_{R2} to the presence of the *nusA* protein. It also has a shorter pause at t_{R2} than wild-type RNA polymerase in the absence of the *nusA* protein (Fig. 4p, r), but does respond normally at t_{R1} to rho factor (not shown). As the *in vitro* binding of the *nusA* protein to the *rif501* RNA polymerase seems to be normal, the polymerase may be unable to recognize *nusA*-dependent terminators. These results imply that, at least at high temperatures, the *nusA* gene protein is an important physiological block to P_R operon expression and λ growth; they also imply that transcription termination *in vitro* by the *nusA* gene protein is an *in vitro* reflection of a phenomenon that occurs *in vivo*. Additional support for the concept that the *nusA* protein causes transcription termination *in vivo* is found in the accompanying article³⁷.

Discussion

We have shown that the *nusA* protein of *E. coli* causes RNA polymerase to pause at a site t_{R2} in the P_R operon of phage λ . Final termination and release of the RNA chain at t_{R2} requires a low concentration of rho factor (a rho factor concentration that is ineffective in the absence of the *nusA* protein) in addition to the *nusA* protein.

How, then, can we reconcile the action of the *nusA* gene protein as a termination factor for RNA synthesis with the demonstration by Friedman²¹ that *nusA* is required to prevent the termination of λ early operon transcription and with the observations of Weissbach and his collaborators that *nusA* is required for the *in vitro* expression of the genes for β -galactosidase²⁹ and the β and β' subunits of RNA polymerase³⁰? We think that there are two major possible explanations.

The first relies on our observation, to be published elsewhere, that the *nusA* gene protein behaves as an RNA polymerase subunit in causing the termination of transcription. It could, then, alter the entire termination specificity of the enzyme so

that, although it enables the enzyme to recognize a new set of termination sites, it also enables the enzyme to ignore another set of termination sites that it would otherwise recognize. In accordance with this principal, it might cause RNA polymerase to read through the rho-dependent terminator in *lacZ* (ref. 31) and the attenuator in front of *rpoB* (ref. 32) even though it causes RNA polymerase to stop at t_{R2} in the P_R operon of λ .

If this idea is correct, it is clearly necessary to repeat, in the presence of the *nusA* gene protein, all the studies done *in vitro* on the effects of rho factor on the transcription of various operons. For example, an interesting anomaly is posed by studies on the attenuator of the *trp* operon of *E. coli*. Termination at *trpA* is efficient *in vitro* in the absence of added factors³⁹ even though genetic studies have suggested that rho factor is involved in termination at *trpA* *in vivo*⁴⁰. It is also particularly interesting to speculate that the insertion element IS1 may possess *nusA*-dependent terminators because it has no rho-dependent or factor-independent terminators *in vitro*⁴¹ and yet is polar in both orientations *in vivo*⁴². In fact, it is already known that the *nusA* protein weakens rho-dependent termination at a terminator in the early region of bacteriophage T7 DNA (M. Chamberlin, personal communication) and that the *nusA* protein strengthens rho-dependent termination at one site and weakens it at other sites at the end of the *trp* operon of *E. coli* (T. Platt, personal communication).

A second intriguing possibility, by no means incompatible with the above, is suggested by our observation that the *N* gene transcription antitermination protein of bacteriophage λ binds directly to the *nusA* gene protein²⁷. This fact suggests how the *nusA* gene protein can be necessary for preventing transcription termination at many sites, including t_{R2} , even though it is a termination factor. We imagine that the λ *N* protein interacts with the RNA polymerase at a *nut* site by binding to the *nusA* gene RNA polymerase subunit. A subtle modification of the *nusA* protein, such as that introduced by the *nusA1* mutation, might block proper binding of the *N* gene protein without significantly interfering with the ability of the modified *nusA* protein to cause transcription termination. The *nusA1* mutant protein, which blocks λ growth *in vivo*, does indeed have termination factor activity *in vitro*. We expect that an inactive *nusA* gene protein would allow λN^- phage to grow and might very well be lethal for the host: these are the properties of a strain carrying the *rif501* RNA polymerase, an RNA polymerase which is resistant to the *nusA* protein *in vitro*.

We can also imagine that *E. coli* might possess operon-specific regulatory proteins, analogous to the *N* gene protein of λ , that bind to the *nusA* gene protein and alter the termination site recognition properties of the RNA polymerase when it is transcribing particular operons. In the case of the *lac* operon, such a regulator might enable RNA polymerase to read through the rho-dependent terminator in the *lacZ* gene and still terminate transcription at the end of the *lac* operon. Such a regulator, if it exists, would probably have been present in the crude *E. coli* extract used in the original L factor assay of the *nusA* protein²⁹. There is no strong genetic evidence for the existence of such a regulator, but some experiments^{43,44} suggest that CAP, the 3', 5'-cyclic AMP-binding protein of *E. coli*, could have such a role. Of course, the existence of such a regulator would not be necessary if the *nusA* gene protein can by itself prevent transcription termination at the rho site in the *lacZ* gene.

This work was supported by the MRC of Canada.

Received 11 December 1980; accepted 6 April 1981.

1. Roberts, J. W. *Nature* **224**, 1158–1174 (1969).
2. Heinemann, S. F. & Spiegelman, W. G. *Cold Spring Harb. Symp. quant. Biol.* **35**, 315–318 (1971).
3. Kourilsky, P., Bourguignon, M.-F. & Gros, F. in *The Bacteriophage Lambda* (ed. Hershey, A. D.) 647–666 (Cold Spring Harbor Laboratory, New York, 1971).
4. Lozeron, H. A., Dahlberg, J. E. & Szybalski, W. *Virology* **71**, 252–277 (1976).
5. Rosenberg, M., Court, D., Shimatake, H. & Brady, C. *Nature* **272**, 414–423 (1978).
6. Skalka, A., Butler, B. & Echols, H. *Proc. natn. Acad. Sci. U.S.A.* **58**, 576–583 (1967).
7. Kumar, S., Bove, K., Guha, A., Hradecna, Z., Maher, V. M. & Szybalski, W. *Nature* **221**, 823–825 (1969).
8. Luzzati, D. *J. molec. Biol.* **49**, 515–519 (1970).
9. Nijkamp, H. J. J., Bove, K. & Szybalski, W. *J. molec. Biol.* **54**, 599–604 (1970).

10. Portier, M.-M., Marcaud, L., Cohen, A. & Gros, F. *Molec. gen. Genet.* **117**, 72-81 (1972).
11. Lozeron, H. A., Aneviski, P. J. & Apirion, D. *J. molec. Biol.* **109**, 359-365 (1977).
12. Court, D. & Sato, K. *Virology* **39**, 348-352 (1969).
13. Hopkins, N. *Virology* **40**, 223-229 (1970).
14. Butler, B. & Echols, H. *Virology* **40**, 212-222 (1970).
15. Salstrom, J. S. & Szybalski, W. *J. molec. Biol.* **124**, 195-221 (1978).
16. Friedman, D. I., Wilgus, G. S. & Mural, R. J. *J. molec. Biol.* **81**, 505-516 (1973).
17. Adhya, S., Gottesman, M. & de Crombrughe, B. *Proc. natn. Acad. Sci. U.S.A.* **71**, 2534-2538 (1974).
18. Franklin, N. C. *J. molec. Biol.* **89**, 33-48 (1974).
19. Segawa, A. & Imamoto, F. *J. molec. Biol.* **87**, 741-754 (1974).
20. Adhya, S. & Gottesman, M. A. *Rev. Biochem.* **47**, 967-996 (1978).
21. Friedman, D. in *The Bacteriophage Lambda* (ed. Hershey, A. D.) 733-738 (Cold Spring Harbor Laboratory, New York, 1971).
22. Georgopoulos, C. in *The Bacteriophage Lambda* (ed. Hershey, A. D.) 639-645 (Cold Spring Harbor Laboratory, New York, 1971).
23. Keppel, F., Georgopoulos, C. & Eisen, H. *Biochimie* **56**, 1503-1509 (1974).
24. Friedman, D., Baumann, M. & Baron, L. S. *Virology* **73**, 119-127 (1976).
25. Friedman, D. I. & Baron, L. S. *Virology* **58**, 141-148 (1974).
26. Friedman, D. I. & Ponce-Campos, R. *J. molec. Biol.* **98**, 537-549 (1975).
27. Greenblatt, J. & Li, J. *J. molec. Biol.* **147**, 11-23 (1981).
28. Greenblatt, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 1991-1994 (1980).
29. Kung, H., Spears, C. & Weissbach, H. *J. biol. Chem.* **250**, 1556-1562 (1975).
30. Zarucki-Schulz, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 6115-6119 (1979).
31. de Crombrughe, B., Adhya, S., Gottesman, M. & Pastan, I. *Nature new Biol.* **241**, 260-264 (1971).
32. Barry, G., Squires, C. L. & Squires, C. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4922-4926 (1979).
33. Greenblatt, J. & Li, J. *Cell* **24**, 421-428 (1981).
34. Roberts, J. W. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3300-3304 (1975).
35. Fianndt, M., Hradecna, Z., Lozeron, H. A. & Szybalski, W. in *The Bacteriophage Lambda* (ed. Hershey, A. D.) 329-354 (Cold Spring Harbor Laboratory, New York, 1971).
36. Georgopoulos, C. P. *Proc. natn. Acad. Sci. U.S.A.* **68**, 2977-2981 (1971).
37. Ward, D. F. & Gottesman, M. E. *Nature* **292**, 212-215 (1981).
38. Lecocq, J.-P. & Dambly, C. *Molec. gen. Genet.* **145**, 53-64 (1976).
39. Lee, F., Squires, C. L., Squires, C. & Yanofsky, C. *J. molec. Biol.* **103**, 383-393 (1976).
40. Korn, L. J. & Yanofsky, C. *J. molec. Biol.* **103**, 395-409 (1976).
41. de Crombrughe, B., Adhya, S., Gottesman, M. & Pastan, I. *Nature new Biol.* **241**, 260-264 (1973).
42. Das, A., Court, M., Gottesman, M. & Adhya, S. in *DNA Insertion Elements, Plasmids, and Episomes* (eds Bukhari, A. I., Shapiro, J. A. & Adhya, S. L.) 93-97 (Cold Spring Harbor Laboratory, New York, 1977).
43. Ullmann, A., Joseph, E. & Danchin, A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3194-3197 (1979).
44. Guidi-Rontani, C., Danchin, A. & Ullmann, A. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5799-5801 (1980).
45. Roberts, T., Shimatake, H., Brady, C. & Rosenberg, M. *Nature* **270**, 274-275 (1977).
46. Schwarz, E., Scherer, G., Hobom, G. & Kössel, H. *Nature* **272**, 410-414 (1978).
47. Lebowitz, P., Weissman, S. & Radding, C. *J. biol. Chem.* **246**, 5120-5139 (1971).
48. Burgess, R. R. & Jendrisak, J. *J. Biochemistry* **14**, 4634-4638 (1975).
49. Weil, P. A. & Blatt, S. P. *Biochemistry* **14**, 1636-1642 (1975).
50. Laemmli, U. K. *Nature* **227**, 680-685 (1970).
51. Burgess, R. R. *J. biol. Chem.* **244**, 6168-6176 (1969).

LETTERS TO NATURE

Microwave emission from flare star AT Mic

O. B. Slee,* I. R. Tuohy,† G. J. Nelson* & C. J. Rennie†

* Division of Radiophysics, CSIRO, PO Box 76, Epping, New South Wales 2121, Australia

† Mount Stromlo and Siding Spring Observatories, Australian National University, Canberra, ACT 2600

There have been several comprehensive optical and low-frequency radio surveys of flare stars^{1,2} but much less attention has been directed to the microwave emission from such stars. Slee and Page³ claimed to have detected one short 5-GHz flare from Proxima Centauri, and several weak, short 5-GHz bursts were reported from YZ CMi (ref. 4) and Proxima Centauri (refs 5, 6) during simultaneous X-ray, optical and microwave observations. We describe here the most convincing detection (7σ) yet made of a microwave flare during simultaneous optical and 5-GHz observations of the dMe 4.5 flare star AT Mic on 25 October 1980. This event was notable for its long duration and the good temporal correlation between the optical and microwave intensities. Unless the dimensions of the source are very much larger than those observed for solar flares, the 5-GHz source has too great a brightness temperature to be produced by the gyrosynchrotron mechanism usually invoked for solar microwave bursts. Much higher energy particles or a coherent emission mechanism must be involved in the stellar microwave flare. The observed correlation between the optical and microwave fluxes from this flare support the theories in which non-thermal electrons are responsible for both emissions.

The optical monitoring of AT Mic was conducted using the 26-inch Yale Columbia refractor at the Mount Stromlo Observatory. A series of 1-min exposures on blue sensitive IIAO plates were taken at intervals of ~5 min. No filter was used, but the response of the emulsion coupled with the UV cutoff of the telescope gives a good approximation to a photoelectric *B* magnitude. Images of AT Mic on the plates were measured with an iris photometer, and the resulting intensity data were normalized relative to a set of comparison stars.

The observing method used in 5-GHz flare observations at Parkes has been described in detail elsewhere³. In the 5-GHz observations high stability against receiver gain fluctuations and thermal emission from passing clouds was achieved by using the equipment in the 'beam wagging' mode. A dual feed at the focus

of the 64-m reflector produces two beams which are alternately placed on and off the flare star, a beam switch and synchronous detector being used to extract the difference signal at 39-Hz between the two beams. A complete integration cycle lasts for 1.7 min, with the driving of the telescope, sampling of the data and analysis of the difference signal being controlled by a computer. The r.m.s. fluctuation level for the integrations is 2.8 mJy.

Figure 1 shows the 5-GHz and optical results for the flare on AT Mic on 25 October 1980. Clearly the microwave emission rose rapidly in less than one integration cycle of 1.7-min duration and thereafter declined slowly over the next 60 min. The base level shown dashed in Fig. 1 was obtained by fitting a straight line to the data taken during an identical observation on the previous evening. The slope evident on this base line is caused by the slow rotation of the aerial pattern on the sky, resulting in small changes in the contributions of confusing sources and ground radiation to the on-source and off-source beam configurations. Some secondary intensity changes may be present on the declining side of the microwave burst but their significance is difficult to assess in view of the relatively high noise level.

The optical exposures give a light curve of similar shape with a peak flare increase corresponding to 0.5-1.0 mag; in view of the significantly higher integration cycle time of ~5 min (as opposed to 1.7 min for the 5-GHz observation), it is difficult to decide whether or not the microwave and optical flares started simultaneously. The peak optical flux is probably considerably

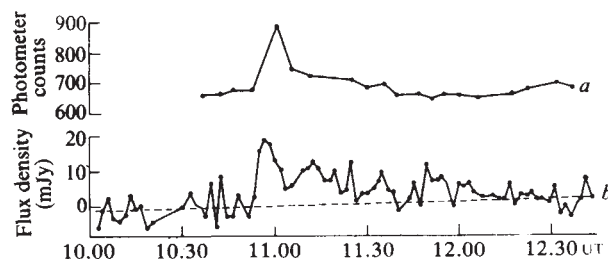


Fig. 1 The optical and 5-GHz light curves for the flare on AT Mic on 25 October 1980. Curve *a* shows a measure of the density of the photographic images, which were exposed for durations of 60 s at intervals of ~5 min. The peak of the optical curve corresponds to a flare luminosity of $\sim 5 \times 10^{30}$ erg s⁻¹. Curve *b* shows the 5-GHz flux densities, which are averaged over 60 s and taken at intervals of 1.7 min. The dashed baseline was obtained by averaging data taken during an identical observation on the previous evening. The r.m.s. fluctuation in the base line is 2.8 mJy.

brighter than shown in Fig. 1 and occurred near the time of the peak 5-GHz flux density.

The photographic measurements enabled the two stellar components of the AT Mic system to be resolved and thus simultaneously monitored. The two stars, separated by ~ 4 arc s, are both believed to be flare-active⁴. Our data unambiguously show that the 25 October flare originated from the southern component of the pair.

There have been previous attempts to observe stellar flares simultaneously at various combinations of X-ray, optical and radio wavelengths^{1-3,5-7}. In fact, only a few flares have been detected simultaneously in more than one of these wavelength ranges. The most striking features of the flare reported here, as compared with the earlier results, are the very long duration and the close correlation between the microwave and optical time profiles.

The 5-GHz flare exhibits a temporal variation similar to that observed in very large solar microwave bursts; for these comparatively rare events the well-known, short-lived impulsive phase is followed by a very much extended second phase which typically lasts for ~ 30 min (ref. 8). Both the impulsive and extended phase of solar microwave bursts are closely associated with hard X-ray emissions^{9,10} and the same population of electrons probably produces both the hard X rays and microwaves^{11,12}. The electrons responsible for the extended phase are generally harder and probably higher in the corona than those generating the impulsive bursts^{8,13}.

It is interesting to consider whether the observed stellar burst can be produced by the synchrotron mechanism which is believed responsible for the solar microwave emissions. In the latter case the synchrotron emission can account for the observed brightness temperatures of up to $\sim 10^9$ K. In the case of AT Mic (assuming a distance of 8.2 pc and a stellar radius of $0.32 R_{\odot}$) the brightness temperature T_b of the 20 mJy flare is $8.5 \times 10^9 / \beta^2$ K, where β is the diameter of the radio source in stellar diameters. Thus if the emission is synchrotron radiation from ~ 100 -keV electrons ($10^8 < T_b < 10^9$ K) the source must have a diameter at least three times larger than the star. If, as is more likely, the source is much smaller than the star, then very much higher brightness temperatures are implied. In this case, particles with energies greater than 100 keV or coherent emission mechanisms need to be considered^{1,2,5}.

Assuming that there are enough high-energy particles in the source region to produce the microwave burst, the same particles (by analogy with the solar flare) may be expected to generate hard X-ray emission with intensity dependent on the ambient (thermal) particle density in the source. Hard X rays have not yet been detected from stellar flares during the several bright optical flares that have occurred during satellite observations in the energy range up to 60 keV (refs 4, 5, 14). On the other hand, soft X rays (0.1–10 keV) have been detected from several flare stars^{6,14,15}, but these are probably of thermal origin and therefore not directly related to the microwave burst.

The observed temporal correlation between the optical and microwave intensities for the flare on AT Mic may only be coincidental. Indeed, if the theory due to Mullan¹⁶ is correct, the optical flare energy is conducted downwards from the coronal plasma which emits the soft X rays. In this case a correlation between the thermal X-ray and optical fluxes would result; any correlation with the non-thermal microwave emission would be much more indirect. On the other hand, if there is a direct relationship between the optical and microwave fluxes, as suggested by the observations in Fig. 1, then theories which are consistent with such a common origin should be investigated. One such theory¹⁷ ascribes the increase in optical brightness to a wavelength shift of IR photons to visible wavelengths by inverse Compton scattering from MeV electrons. The mechanism may be very efficient in the cool flare stars, which have most of their quiescent emission in the IR region of the spectrum. This theory has recently been supported by Bruevich *et al.*¹⁸, who observed decreases in the IR fluxes of several flare stars immediately before flares in the visible spectrum; the magnitudes and dura-

tions of the negative IR flares were sufficient to account for the excess flare energy radiated subsequently in the visible spectrum. The significance of this mechanism for the current observations is that the same MeV electrons responsible for the optical emission could also produce the microwave flare, thus accounting for the observed correlation.

The fact that the present 5-GHz flare on AT Mic lasts much longer than the microwave bursts generally observed either from the Sun or from the flare stars Proxima Centauri^{5,6} and YZ CMi (ref. 4) suggests that the extended part, at least, may result from a previously unobserved mechanism. On the other hand, long-duration optical flares¹⁷ and metre-wave radio bursts² have been previously observed on flare stars. It may simply be that the extended parts of previous stellar microwave bursts have been below the threshold for detection.

We thank Mr I. Wilson for assistance in the 5-GHz observations. The radio and optical observations of AT Mic were prompted by K. Mason and S. Kahn in support of a planned simultaneous X-ray observation by the Einstein Observatory (which unfortunately did not occur).

Received 12 January; accepted 8 May 1981.

- Spangler, S. R. & Moffett, T. J. *Astrophys. J.* **203**, 497 (1976).
- Nelson, G. J. *et al. Mon. Not. R. astr. Soc.* **187**, 405 (1979).
- Slee, O. B. & Page, A. A. *Proc. IAU Colloq.* **46**, 150 (1979).
- Kunkel, W. E. *Proc. IAU Symp.* **67**, 15 (1975).
- Karpen, J. T. *et al. Astrophys. J.* **216**, 479 (1977).
- Haisch, B. M. *Astrophys. J. Lett.* **225**, L35 (1978).
- Haisch, B. M. *et al. Astrophys. J.* **245**, 1009 (1981).
- Frost, K. J. & Dennis, B. R. *Astrophys. J.* **165**, 655 (1971).
- Crannell, C. J., Frost, K. J., Mätzler, C., Ohki, K. & Saba, J. L. *Astrophys. J.* **223**, 620 (1978).
- Stewart, R. S. & Nelson, G. J. *Proc. astr. Soc. Aust.* **3**, 390 (1979).
- Nelson, G. J. & Stewart, R. T. *Proc. astr. Soc. Aust.* **3**, 392 (1979).
- Holt, S. S. & Ramaty, R. *Solar Phys.* **8**, 119 (1969).
- Hoynig, P., Brown, J. C. & van Beek, H. F. *Solar Phys.* **48**, 197 (1976).
- Kahn, S. M. *et al. Astrophys. J. Lett.* **234**, L107 (1979).
- Heise, J. *et al. Astrophys. J. Lett.* **202**, L73 (1975).
- Mullan, D. J. *Astrophys. J.* **207**, 289 (1976).
- Gurzadyan, G. A. *Flare Stars* (Pergamon, Oxford, 1980).
- Bruevich, V. V. *et al. Proc. Crimean Astrophys. Obs.* **61**, 90 (1980).

Fine structure in the ionospheric D-region

E. V. Thrane & B. Grandal

Norwegian Defence Research Establishment, PO Box 25, N-2007 Kjeller, Norway

T. Flå & A. Brekke

Institute for Mathematical and Physical Sciences, University of Tromsø, N-9001 Tromsø, Norway

Assuming that their observations of the weak echoes of high frequency radio waves scattered from the lower ionosphere were caused by Fresnel reflections from discontinuities in the refractive index, Gardner and Pawsey¹ developed a method for determining the height distribution $N_e(h)$ of the D-region electron densities from observations of the ratio $A_x/A_0(h)$ of the back-scattered amplitudes of the extraordinary and ordinary magneto-ionic components. This 'partial reflection' method turned out to be a useful tool for studies of the ionospheric D-region²⁻⁵. We compare here observations of weak radio echoes from the ionospheric D-region, with model computations of radio wave scattering, based on *in situ* measurements of ionospheric irregularities. The results show excellent agreement, indicating that an important source of the weak partial reflections has been identified.

The nature of the irregularities and the scattering mechanism has been extensively discussed and the theory has been gradually developed to describe different types of scattering⁶⁻¹³. Part of this discussion dealt with the possibility that the irregularities

in refractive index could be due to changes in electron-neutral collision frequency⁷ as well as changes in electron density, as originally assumed by Gardner and Pawsey. Belrose *et al.*¹⁰, however, presented data supporting the original hypothesis, and in the following we assume that only changes in electron density cause the partial reflections. The remaining important problem is to map the structure of these electron density irregularities. Because the air in the D-region is only weakly ionized, the degree of ionization being in general smaller than 0.1 p.p.m., we may expect a close coupling between changes in the ionized and neutral constituents. Studies of irregularities in ionization density may therefore give information about the dynamical state of the mesosphere.

The simple theory of Fresnel reflection implicitly assumes that sharp vertical gradients in refractive index exist extending horizontally over at least a Fresnel zone (≈ 5 km for HF-waves near 70-km altitude). The vertical change in electron density $\Delta N_e/N_e$ must be of the order of 1–8% within a fraction of a radio wavelength (30–100 m) to produce the observed scattered amplitudes¹⁴. A revised theory by Flood⁸ assumes volume scattering from a medium in which irregularities with a spectrum of sizes are present. Constructive interference of backscattered signals will occur where the spatial variations have scales near half the radio wavelength. The radar will therefore preferentially be sensitive to the intensity of fluctuations with such scales. Ground based experiments to determine the nature of the scattering mechanism have not yielded conclusive results, but indicate that both Fresnel reflection from extended irregularities and volume scattering may occur. There seems to be a tendency for the larger irregularities to occur at low heights (below 70 km)^{14,15}.

We now compare observations of partial reflections from the D-region with computed intensities of radio echoes. The model computations are based on volume scattering from irregular ionospheric structures measured *in situ* in a sounding rocket. The experimental data were obtained from a rocket-borne electrostatic ion-current probe combined with simultaneous recordings of partial reflections. The experiments were made in North Norway, at the Andøya Rocket Range (69°17'N, 16°01'E) and at Ramfjord Research Station (69°58'N, 19.22°E) where the University of Tromsø operates a partial reflection facility. The latter station is situated near Tromsø, 120 km to the north-east of the Andøya Rocket Range.

On 1 March 1978 at 01.13UT a research rocket code F-47, was launched from Andøya Rocket Range into a moderately disturbed auroral ionosphere. Before and during launch the ionospheric conditions were monitored by the partial reflection experiment at Ramfjord and by riometers and magnetometers at both sites. The launch criteria were: the presence of strong, stable partial reflections in the height range 65–95 km, and that both riometer absorption and magnetic activity were of similar magnitudes at the two sites. At the time of launch a moderate disturbance with riometer absorption between 0.5 and 1 dB at both sites had lasted for several hours with only small changes in absorption and partial echo structure. Magnetic activity was low during this period.

The F-47 payload was carried by a Nike Apache rocket and reached an apogee of 127.5 km. The electrostatic probe was designed to measure positive ion current with high time resolution and accuracy. The sampling rate was 1,136 Hz, providing one measurement of ion current per metre along the rocket trajectory through the D-region. The rocket experiment and the results have been described elsewhere¹⁶. Some important assumptions made in the analysis of the rocket data should be mentioned. First, the observed ion current I is assumed to be proportional to ambient ionospheric ion number density N_i over small height intervals ($\Delta h \approx 1$ km). As the probe moves through the plasma with a speed about three times larger than the thermal velocity of the ions, the observed time variation in ion current is assumed to reflect the spatial structure of irregularities in the ambient ion density, that is $\delta I/I = \delta N_i/\langle N_i \rangle$, where δI and δN_i are deviations from the respective running means $\langle I \rangle$

and $\langle N_i \rangle$ taken along a suitable length of the rocket trajectory. Furthermore, local charge neutrality is assumed so that $\delta N_i/N_i = \delta N_e/N_e$ where N_e is the electron number density. The instrument provides power spectra in the scale range 3–100 m. In the D-region the spectral slope is near $-5/3$, indicating that the irregularities are caused by homogeneous, isotropic turbulence, whereas the spectra observed above 95 km have a 'white noise' character. Of particular interest is the intensity in the D-region of irregularities with scales equal to $1/2 \lambda$ where λ is the radio wavelength of the partial reflection radar. The height variation of this intensity is shown in Fig. 1. The accuracy of the rocket results shown in Fig. 1 depend on several factors. The determination of the relative fluctuations in positive ion density $\sqrt{\langle \delta N_i^2 \rangle} / \langle N_i \rangle$ at the length scale $\lambda/2$ is accurate to better than 5%. The absolute magnitude of the electron density fluctuations is found from the equation

$$\sqrt{\langle \delta N_e^2 \rangle} = \frac{\sqrt{\langle \delta N_i^2 \rangle}}{\langle N_i \rangle} \langle N_e \rangle$$

where the electron density profile $\langle N_e \rangle(h)$ is determined from a radio wave propagation experiment in the rocket (Faraday rotation and differential absorption measurements, M. Friedrich personal communication). The structure in Fig. 1 represents a real variation in the fluctuation intensity, it is not associated with any large-scale structure in the electron density profile $\langle N_e \rangle(h)$, which is a smooth monotonic function of height¹⁶. The absolute accuracy of this profile changes with height from $\sim 30\%$ near 65 km to better than 10% at 95 km. Note that the observed fluctuations in ion density could not have been caused by fast time variations in the flux of the ionizing source (precipitating energetic electrons), because the time constant of the ions in the lower ionosphere is of the order of 100–1,000 s.

The partial reflection experiment (PRE)^{17,18} facility near Tromsø operates on 2.75 MHz ($\lambda = 109$ m) and using a 17 dB gain antenna, transmits circularly polarized, 20 μ s wide, pulses vertically with a pulse repetition rate of 50 Hz. The equivalent

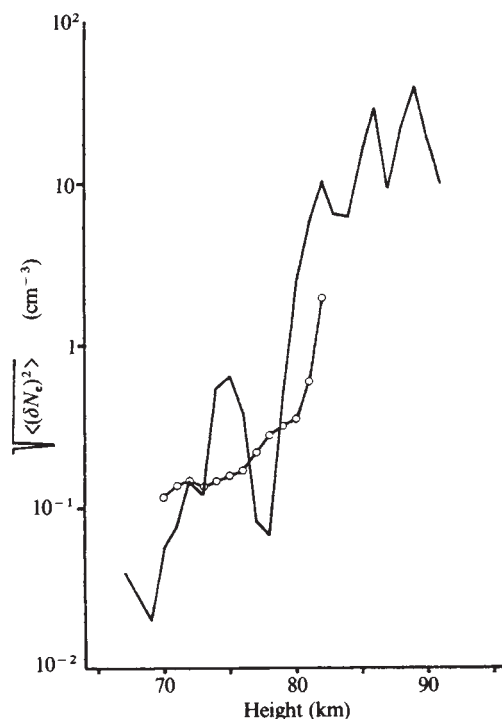


Fig. 1 Root mean square fluctuation of ion density for scales of $\frac{1}{2}\lambda$ ($\lambda = 109$ m, the radio wavelength for the PRE radar). —, Fluctuations derived from the rocket measurements; -○-, fluctuations computed from the PRE measurements, using the amplitude of the ordinary magneto-ionic component.

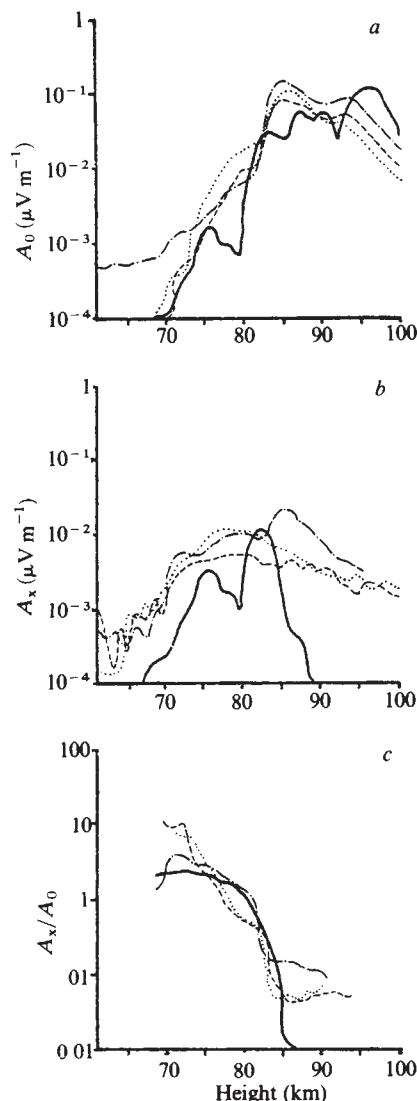


Fig. 2 Comparison between partial reflection measurements and results from calculations of partial reflection amplitudes based on rocket measurements of ionization density fluctuations. —, Computed values; ----, measurements 01.12.23–01.13.09 UT (1 March 1978); — · —, measurements 01.13.54–01.14.34 UT; · · ·, measurements 01.15.21 UT–01.16.02 UT. *a*, r.m.s. amplitude of ordinary magneto-ionic component versus height; *b*, r.m.s. amplitude of extraordinary magneto-ionic component versus height; *c*, r.m.s. amplitude ratio A_x/A_0 versus height.

isotropically radiated peak power is about 125 kW. The polarization is switched between pulses, so that alternate pulses correspond to the ordinary and extraordinary magneto-ionic modes. The transmitter and receiver systems are computer controlled, and the weak partial echoes are recorded on magnetic tape for every 1-km height interval, typically in the height region 50–100 km. Three examples of PRE observations of ordinary echo amplitude $A_0(h)$ and extraordinary amplitude $A_x(h)$, averaged over 40 s intervals during the rocket flight are shown in Fig. 2. The electron density profile $N_e(h)$ may be derived from the ratio $A_x/A_0(h)$, and this ratio is also shown. Observations from the three different time intervals are quite consistent, indicating that no significant ionospheric changes occurred during the flight. The height distribution of electron number density was also determined from Faraday rotation measurements in the rocket (M. Friedrich, personal communication). These number densities agreed well with those determined from the PRE results. Thus the D-region showed no significant changes over a distance of 120 km.

The ratio of received to transmitted powers for the partial reflections may be expressed as⁸ (T. F. unpublished data):

$$\frac{P_R(h)}{P_T} = \text{constant} \frac{1}{h^2} \int_0^\infty dz_1 g^2\left(\frac{z_1-h}{2}\right) |C(z_1)|^2 \times \exp\left(-4 \int_0^{z_1} k_i(z') dz'\right) \tilde{R}(z_1) \quad (1)$$

Here h is the height from which the power $P_R(h)$ is received, $g((z_1-h)/2)$ represents the pulse shape, $C(z_1)$ is a known function based on an assumed height variation of electron-neutral collision frequency¹⁹, and k_i is the imaginary part of the wavenumber. The constant is determined by calibration of the radar system, and $\tilde{R}(z_1)$ is the intensity of ionospheric fluctuations with scales corresponding to $\lambda/2$ or $2k_r$, where k_r is the real part of the wavenumber.

The relation between the rocket and partial reflection observations was tested in two ways. First, the amplitudes $A_x(h)$, $A_0(h)$ and ratio $A_x/A_0(h)$ were computed directly from equation (1), using values of $\tilde{R}(z_1)$ determined from the rocket probe for 1-km intervals. In Fig. 2 the results of the computations are compared with the measurements. The agreement, both regarding absolute magnitudes and height variation is very good. The main sources of uncertainty are the possible ionospheric changes over the distance of 120 km between the radar and the rocket, and possible inaccuracies in the absolute calibration of the radar system. The discrepancy between calculated and measured amplitudes for the extraordinary wave above 85 km, could be due to contamination of the measurements by oblique echoes, or to the effects of not correcting for the change in the extraordinary mode group velocity above 85 km.

In the second test equation (1) was solved for the term $\tilde{R}(z_1)$, using the measured ratio $P_R(h)/P_T$. To achieve this, the pulse $g((z_1-h)/2)$ was assumed to be a square wave of width 3 km centred at a height h , and all other terms inside the integral were assumed constant within this height range. In Fig. 1 the computed term $\tilde{R}(z_1)$ representing the intensity of the fluctuations is compared with the rocket measurements. Again the agreement is good, although the computations do not reproduce the detailed structure observed in the rocket. At least part of this discrepancy could be due to the fact that the computed fluctuation intensity is averaged over a scattering volume of width 3 km.

We conclude that the intensity and height variation of partial reflections measured during the rocket flight are consistent with volume scattering from the ionospheric irregularities measured by the electrostatic probe in the rocket. The power spectra of ion density fluctuations derived from the rocket data, indicate that these fluctuations are caused by turbulence in the neutral air¹⁶. We may therefore conclude that such turbulence is an important source of partial reflections.

Received 23 February; accepted 19 May 1981.

- Gardner, F. F. & Pawsey, J. L. *J. atmos. terr. Phys.* **3**, 321–344 (1953).
- Fejer, J. A. & Vice, R. W. *J. atmos. terr. Phys.* **16**, 291 (1959).
- Gregory, J. B. *J. geophys. Res.* **66**, 429–435 (1961).
- Belrose, J. B. & Burke, M. J. *J. geophys. Res.* **69**, 2799 (1962).
- Thrane, E. V., Haug, A., Bjelland, B., Anastassiades, M. & Tsagakis, E. *J. atmos. terr. Phys.* **30**, 135–150 (1968).
- Gallett, R. M. *Proc. IRE* **43**, 1240–1252 (1955).
- Piggott, W. R. & Thrane, E. V. *J. atmos. terr. Phys.* **28**, 311–314 (1966).
- Flood, W. A. *J. geophys. Res.* **73**, 5585–5598 (1968).
- Austin, G. L. & Manson, A. H. *Radio Sci.* **4**, 35–40 (1969).
- Belrose, J. S., Burke, M. J., Coyne, T. N. R. & Reed, J. E. *J. geophys. Res.* **77**, 4829–4838 (1972).
- Briggs, B. H. & Vincent, R. A. *Aust. J. Phys.* **26**, 805 (1973).
- Rastogi, P. K. & Bowhill, S. A. *J. atmos. terr. Phys.* **38**, 399 (1976).
- Hocking, W. K. *J. geophys. Res.* **84**, 845–851 (1979).
- Vincent, R. A. & Belrose, J. S. *J. atmos. terr. Phys.* **40**, 35–47 (1978).
- Schlegel, K., Brekke, A. & Haug, A. *J. atmos. terr. Phys.* **40**, 205–213 (1978).
- Thrane, E. V. & Grandal, B. *J. atmos. terr. Phys.* **43**, 179–189 (1981).
- Haug, A., Thrane, E. V., Bjørnå, K., Brekke, A. & Holt, O. *J. atmos. terr. Phys.* **39**, 1333–1340 (1977).
- Holt, O., Brekke, A. & Hansen, T. *Proc. 5th ESA Symp.*, ESA SP-152, 387–392 (1980).
- Thrane, E. V. & Piggott, W. R. *J. atmos. terr. Phys.* **28**, 721–737 (1966).

Photovoltaic conversion by macromolecular thionine films

R. Tamilarasan & P. Natarajan*

Department of Chemistry, University of Madras,
Tiruchirapalli 620 020, Tamil Nadu, India

On excitation by light, thionine dye oxidizes Fe^{2+} ions in solution and the photoproducts undergo dark reactions to restore the starting materials. The photogalvanic (photovoltaic) potentials generated by this cyclic process can be demonstrated by means of a cell with two platinum electrodes, one illuminated and one in the dark. Light is converted more effectively into electricity in the totally illuminated thin-layer (TITL) cell with platinum and tin oxide electrodes, but the efficiency of the conversion is limited by energy-wasting back reactions in solution and by restriction on the concentration of the dye¹. We now show that these limitations can be overcome if thick films ($\sim 10\ \mu\text{m}$) of thionine condensed with macromolecules are coated on to inert electrodes.

Recent investigations of photoredox reactions in heterogeneous media and of charge-transfer reactions in micellar systems², organized monolayers³ and chemically modified electrode systems⁴ are aimed at understanding the reactions involving biological membranes including photosynthetic apparatus. Microheterogeneous environments drastically change the features of chemical reactions occurring in homogeneous media. In a photoinduced electron-transfer reaction, the products readily recombine in a homogeneous medium to regenerate the starting materials whereas charge separation is favoured in the micellar surface or in organized monolayer assemblies⁵. Such systems could be used to store and utilize light energy. Here we describe the photocatalysed redox reactions induced by the thionine covalently bound to poly(*N*-methylolacrylamide) and related copolymers. Films of polymer-bound dye were coated on to inert electrodes. The electrode reactions are totally different from those of the excited dye in homogeneous solution.

Purified thionine on condensation⁶ with poly(*N*-methylolacrylamide) produces the polymer-dye complex, poly(acrylamidomethylthionine-co-methylolacrylamide). Uncondensed dye is removed by dialysing the reaction mixture for several days with water. The ratio of thionine units present to the hydroxymethylacrylamide units in the macromolecular chain could be varied depending on the reaction conditions; the monomer/dye (M/D) ratio in the polymer-dye complex is 120 ± 10 . In solution the dye centres present in the polymeric chain do not form aggregates at this M/D ratio. However, with increasing amounts of dye present in the chain, the dye centres aggregate and blueshift the absorption maximum by 15 nm with respect to the free dye⁷. Evaporation of the solvent in a flash evaporator at 40 °C leaves behind an insoluble dye-polymer complex: presumably some cross-linking occurs. However, the polymer-dye complex remains in solution without forming a precipitate when the solvent is not evaporated. A drop of the polymer-dye solution placed on a glass or metal surface produces a film on evaporation in vacuum at 90 °C. The absorption spectra of the polymer-dye in solution and of the film coated on to an electrode plate are shown in Fig. 1. In the film some of the dye has aggregated as indicated by the absorption spectrum. In all the present experiments a 1-cm² platinum electrode was coated with the polymer-dye. The film was 12–15 μm thick.

The photoelectrochemical cell consisted of a platinum electrode covered with a film of the polymer-thionine complex, and a plain platinum electrode. The electrodes were immersed in a de-aerated solution containing 10^{-2} M ferrous sulphate in 0.5×10^{-2} M sulphuric acid. The dye-coated electrode was irradiated

Table 1 Polymer-dye coated electrodic behaviour on irradiation

Substrate	ΔE_{oc} (mV)	I_{sc} (μA)	P_{max} (μW)
A	+37.4*	2.84	0.027
	+31.4†	3.70	0.029
B	+32.0*	1.80	0.014
	+29.6†	2.76	0.020
C	+ 4*	—	—
	+ 3†	—	—

A, Poly(*N*-acrylamidomethylthionine-co-methylolacrylamide); B, poly(*N*-acrylamidomethylthionine-co-methylolacrylamide-co-acrylic acid); C, poly(*N*-acrylamidomethylthionine-co-methylolacrylamide-co-vinyl pyridine).

* Stationary condition; † stirring condition.

by a 300 W tungsten projection lamp. The open-circuit photopotential (ΔE_{oc}), short circuit current (I_{sc}) and maximum power output (P_{max}) were measured from the plots of photopotential, photocurrent and power output as a function of applied resistance. The results are shown in Table 1 for the thionine bound to poly(*N*-methylolacrylamide) and its copolymers. For comparison, the power output and the potential generated on irradiation in identical conditions in the TITL⁸ cell with the distance of 1 mm between the SnO_2 and platinum electrode were measured for unbound thionine. The open-circuit potential measured was -74.5 mV and the power output $0.038\ \mu\text{W}$ for unstirred solutions. The direction of the current is reversed with respect to the plain platinum electrode in the TITL cell compared with that with polymer-dye film coated electrode.

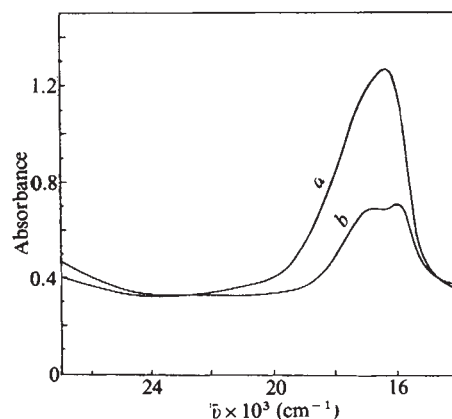


Fig. 1 Absorption spectra of poly(*N*-acrylamidomethylthionine-co-methylolacrylamide) in aqueous solution (a) and as a film (b).

For the electrode reactions occurring at the dye-coated electrode current-voltage curves were determined in a PAR cyclic voltammeter. The results (Fig. 2) indicate that the thionine-leucothionine couple is more reversible at the dye coated electrode than the $\text{Fe}^{2+}/\text{Fe}^{3+}$ couple. The peak potentials of the polymer-bound thionine/polymer-bound leucothionine are not affected by coating it on to a platinum electrode. However, the cathodic peak potentials of $\text{Fe}^{2+}/\text{Fe}^{3+}$ system is shifted cathodically by 110 mV at the polymer coated electrode compared with the uncoated platinum electrode. Albery *et al.*⁹ have reported that electrolytically coated thionine on a platinum electrode shows a similar behaviour.

That the polymer-dye coated electrode functions as a cathode on irradiation suggests that the electrode reaction reduces Fe^{3+} or some other species. However, the reduction of free Fe^{3+} ions is not a favourable process at this electrode. We propose that the electroactive species is a complex of semithionine with Fe^{3+} ions surrounded by the polymer network along with solvent. As thionine is immobilized in a polymer film the photoproduct semithionine does not undergo the disproportionation reaction

* To whom correspondence should be addressed.

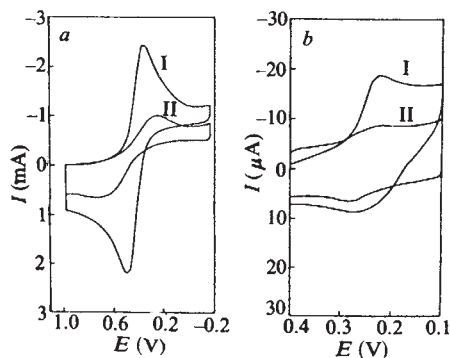
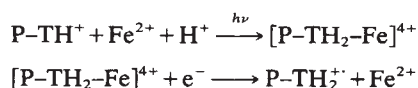


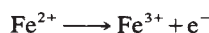
Fig. 2 Cyclic voltammograms versus SCE. Scan rate, 100 mV s^{-1} , electrode area, 1 cm^2 , in $0.25 \text{ M } [\text{H}_2\text{SO}_4]$. a, $\text{Fe}^{2+}/\text{Fe}^{3+}$ couple at (I) plain platinum and (II) polymer-thionine coated electrodes (10^{-2} M of Fe^{2+}). b, Polymer-bound thionine/leucothionine couple at (I) plain platinum electrode with polymer-dye in solution ($5 \times 10^{-5} \text{ M}$ dye); and (II) for the polymer-dye coated electrode without the polymer-dye in solution.

observed in homogeneous medium. The electrode reactions are thus:

Cathode reaction



Anode reaction



Fe^{3+} ions produced at the anode combine with the semithionine at the cathode to generate the starting materials. The proposed mechanism implies that the oxidation of the semithionine in the complex $[\text{P-TH}_2\text{-Fe}]^{4+}$ present near the electrode seems to be a much slower process than the reduction of the complex at the electrode. Oxidation of P-TH_2^+ by the Fe^{3+} ions produced at the anode probably occurs by charge migration through interacting thionine-semithionine centres in the polymer network. Charge can be transferred in thin films of dyes and polymers coated on to electrodes and electrons can tunnel through stacks of monolayers¹⁰. However, in quite thick films of polymer a high ohmic resistance is expected to result in negligible amount of current flow. In the present case with the $\sim 15\text{-}\mu\text{m}$ thick polymer film, charge transfer to electrode must be explained either by the migration of charge through the dye centres or the penetration of ferrous ion through the porous polymer network or a combination of both. It has been suggested¹¹ that the polymer films of $10 \mu\text{m}$ thickness in suitable solvents exhibit considerable flexibility, which facilitates charge transfer through bound chromophores.

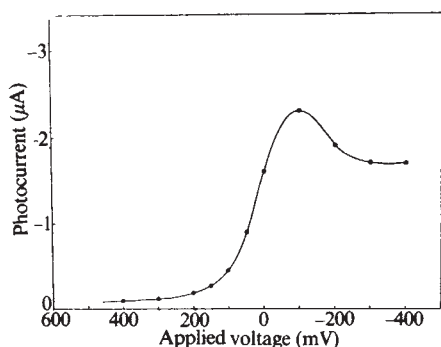


Fig. 3 Photocurrent plotted against applied voltage (versus SCE) across the working electrode-reference electrode for the polymer-dye coated electrode.

Decrease in the open-circuit photopotential and the practically negligible amount of current flow observed in the case of poly(acrylamidomethylthionine-co-methylolacrylamide-co-vinylpyridine) indicates the inhibition of free movement of Fe^{2+} ions through the polymer network due to the strong binding between the ferrous ion and pyridine ligand. With the corresponding polyacrylic acid copolymer at pH 2.0, as much of the carboxylic acid is in the protonated form, lack of coordination of acrylate with Fe^{2+} ions does not affect the mobility of these ions through the polymer network.

The polymer-dye coated electrode in the electrochemical cell was illuminated at various applied potentials and the resulting photocurrents are shown in Fig. 3.

The present study has shown that high concentrations of the light absorbers in contact with electrodes facilitate absorption of more light per unit thickness and bring about specific electrode reaction at the electrode. The catalyst, which brings about the light induced electrode reaction, is mounted on the electrode, eliminating unwanted waste-ridden bulk reactions. Comparison of the TITL cells, which show the maximum efficiency for the Fe^{2+} -thionine photogalvanic cell, with the polymer-dye coated electrode system shows that nearly 75% of the power output of the former cell is attained by absorbing light in less than one-hundredth of the thickness of the thin layer. As the bulk solution contains no dye, operation of the system becomes simpler. The polymer-dye film is stable for days in the dark and under illumination shows no observable change in its electrochemical behaviour.

This research is supported by grants from University Grants Commission and by the Department of Science and Technology.

Received 20 March; accepted 1 May 1981.

1. Albery, W. J., Bowen, W. R., Archer, M. D. & Ferreira, M. I. *J. Photochem.* **11**, 27-37 (1979).
2. Gratzel, M. *Isr. J. Chem.* **18**, 364-368 (1979).
3. Kuhn, H. *Pure appl. Chem.* **61**, 341-352 (1979).
4. Wrighton, M. S. *Chem. Engng News* 29-47 (1979).
5. Whitten, D. G. *Accs Chem. Res.* **13**, 83-90 (1980).
6. Kamogawa, H., Kato, M. & Sugiyama, H. *J. polym. Sci. A1* **6**, 2967-2991 (1968).
7. Tamilarasan, R. & Natarajan, P. *Ind. J. Chem.* **20A**, 213-215 (1981).
8. Hall, D. E., Wildes, P. D. & Lichtin, N. N. *J. electrochem. Soc.* **125**, 1365-1371 (1978).
9. Albery, W. J. *et al. Nature* **282**, 793-797 (1979).
10. Oyama, N. & Anson, F. C. *J. electrochem. Soc.* **127**, 640-647 (1980).
11. Kaufmann, F. B. & Engler, E. M. *J. Am. chem. Soc.* **101**, 547-549 (1979).

Viscosity of high-pressure ice VI and evolution and dynamics of Ganymede

J. P. Poirier, C. Sotin & J. Peyronneau

Institut de Physique du Globe, Université Paris VI, 4, Place Jussieu, 75320 Paris Cedex 05, France

Current models¹⁻⁵ of the evolution of the jovian satellite, Ganymede, assume that it was formed by homogeneous accretion of water ice and silicate particles (about 50 mass%) with chondritic abundance of the radioactive elements U, Th and K. As the satellite reached its final size it would probably have been composed of a mixture of silicate particles and high-pressure phases of ice, from the centre outwards: ice VIII, ice VI, ice II and finally, near the surface, ice I⁴. Due to radioactive decay, the temperature would rise in the interior, presumably transforming ice VIII into ice VII, and bringing ice VII and VI close to their melting point. We have measured the viscosity of the high-pressure ice VI at room temperature and pressures of 1.1-1.2 GPa in a sapphire anvil cell; fine particles were used to visualize the flow of ice down the radial pressure gradient. The low value we found for the viscosity ($\eta \approx 10^{14} \text{ P}$) suggests that solid state convection might have taken place during the early evolution of Ganymede, thus preventing melting and differentiation.

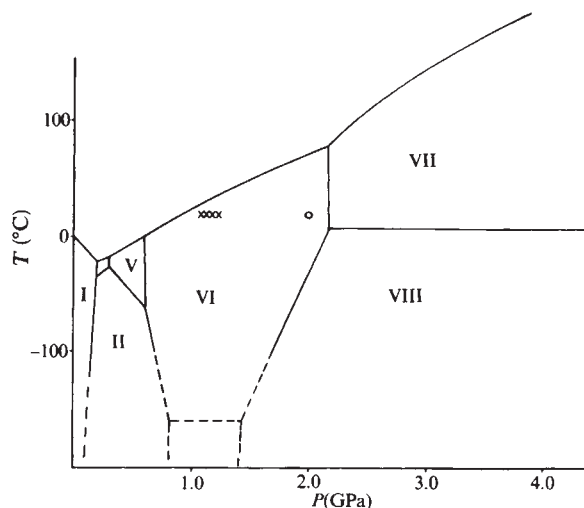


Fig. 1 Phase diagram of ice⁶ showing our experimental conditions (×); and conditions for which the value of viscosity was extrapolated (○).

Parmentier and Head³ have shown that if the viscosity of high-pressure ices near their melting point is lower than 10^{17} P, solid state convection can take place, effectively removing heat and preventing melting and differentiation of the planet. However, if the viscosity is higher than 10^{17} P, melting could eventually occur leading to differentiation and silicates could sink forming a rocky core surrounded by a liquid water mantle and a lithosphere of ice I and II¹; if the viscosity of ice I is low enough ($\approx 10^{14}$ P) the lithosphere itself can convect², extract heat from the liquid layer which will freeze in a time of the order of 10^8 yr, and form a mantle composed mostly of ice VI⁴ which, in turn, can convect if its viscosity is low enough. More generally, if the viscosity of ice in the interior were as low as 10^{14} P, a steady state body, either differentiated or undifferentiated, would be primarily solid⁵.

Ice VI is easily formed in a diamond anvil cell and observed by optical microscopy⁶. The phase diagram of ice⁷ (Fig. 1) shows that liquid water transforms to ice VI at 0.9 GPa, at 20 °C. For such relatively low pressures, the anvils can be made of sapphire instead of diamond with the result that a gasket with a larger hole can be used, allowing a wider field of observation. The viscosity of the ice can be measured by following the motion of marker particles down the radial pressure gradient in the cell.

A cell of the usual Block–Piermarini type was fitted with anvils machined from a 6-mm diameter single crystalline rod of Al_2O_3 with the *c*-axis along its length. The anvils were 5 mm

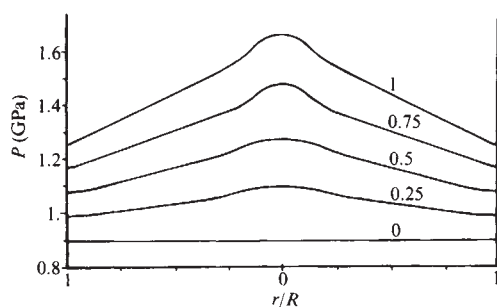


Fig. 2 Hydrostatic pressure in the sapphire anvil cell plotted against r/R , where r is the distance from the centre and R the radius of the hole in the gasket. Each curve corresponds to a different load and is labelled in fractions of one turn of the screw.

high and were tapered to an end diameter of 1.3 mm. The gaskets were made of annealed copper foil 0.35 mm thick with a 0.6-mm diameter hole. The hole is filled with tap water containing particles of 2 μm alumina grit and the load is applied, through a lever, by turning a screw (by capillarity, most of the particles stay close to the anvil surfaces); the cell is installed under an optical microscope. At first the gasket is extruded a little inwards, reducing the diameter of the hole and the hydrostatic pressure increases until 0.9 GPa, when ice VI crystals begin to form, and eventually fill the hole. From then on, further increases in the load cause the ice to be extruded outwards and the diameter of the hole to increase slightly. There is a radial pressure gradient from the maximum pressure at the centre to the edge and it can be reasonably assumed that the ice flows fully plastically, having reached its shear strength. In such cases, one can assume that the radial pressure gradient is equilibrated by the shear stresses along the anvil surfaces⁸ and that

$$\frac{\delta P(r)}{\delta r} = \frac{2\sigma_{rz}}{h} \quad (1)$$

where $P(r)$ is the hydrostatic pressure at a distance r from the centre, σ_{rz} is the shear stress along the anvil surfaces (in cylindrical coordinates); h is the thickness of the ice disk which can be measured for a given load by focusing the microscope on the

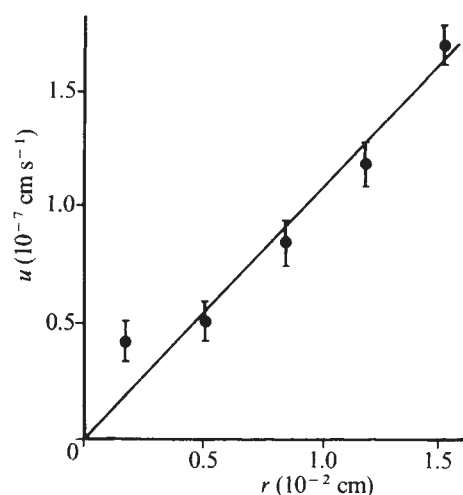


Fig. 3 Plot of the radial velocity (u) of the markers against radius (r).

upper and lower surfaces. $P(r)$ is determined for several values of the load (expressed by N , number of $\frac{1}{8}$ turns of the screw after the ice has formed) by using fixed transition points for alkali halides at room temperature: 0.36 and 1.16 GPa for NH_4F (ref. 9), 0.52 GPa for RbCl (ref. 10) and 1.8 GPa for KBr (ref. 11). Holes in gaskets 150- μm thick (the thickness of the gasket when the ice starts to form) are filled with alkali halide powder. As the load is increased the powder first sinters and the transformation appears at the centre of the hole; the transformed phases can be seen as dark circular patches propagating outwards as the load is increased. The pressure at the value of r corresponding to the rim of the dark patch is equal to the transformation pressure. For each of several values of r , it is possible to determine the number of $\frac{1}{8}$ turns of the screw for which the transitions appear; these data are converted into smoothed curves giving $P(r)$ for $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$, 1 turn of the screw. We assume that the same curves will be approximately correct for ice VI if we take $P(r)$ to be the pressure over 9 kbar for the same number of turns of the screw after formation of the ice. The pressure distribution in ice VI is shown in Fig. 2.

Table 1 Experimental results at various distances from the centre r

r/R	r (cm)	u (cm s ⁻¹)	$\dot{\gamma}_{rz}$ (s ⁻¹)	P (GPa)	$\delta P/\delta r$ (cgs)	η (P)	$T_M - T$ (K)
0.25	4.2×10^{-3}	4.2×10^{-8}	3.2×10^{-6}	1.22	1.4×10^{11}	1.4×10^{14}	16
0.5	8.4×10^{-3}	8.4×10^{-8}	6.5×10^{-6}	1.17	1.2×10^{11}	6.1×10^{13}	14
0.75	1.3×10^{-2}	1.3×10^{-7}	10^{-5}	1.12	1.2×10^{11}	4×10^{13}	12
1	1.7×10^{-2}	1.7×10^{-7}	1.3×10^{-5}	1.08	9.5×10^{10}	2.4×10^{13}	10

u is the radial velocity of the particles; $\dot{\gamma}$, the shear strain rate; P , the hydrostatic pressure; $\delta P/\delta r$, the pressure gradient; η , the viscosity; $T_M - T$ is the distance to the melting point.

The viscosity has been determined for the load corresponding to $\frac{1}{2}$ turn of the screw after formation of the ice. Micro-photographs of the field of ice with particles of alumina close to the anvil surfaces were taken 2 and 17 min after the load was applied. Comparison of enlarged photographs shows that the particles at a distance r from the centre of the cell have moved radially by Δr during the time Δt between the two photographs. Their average velocity $u(r) = \Delta r(r)/\Delta t$ is found to be proportional to r : $u(r)/r \approx 10^{-5} \text{ s}^{-1}$ (Fig. 3). The markers are passively carried by the outwards creeping ice. As a first approximation, we take $\dot{\gamma} = u(r)/h$ for the viscous shear rate of ice.

$\dot{\gamma}$ is related to the shear stress by

$$\sigma_{rz}(r) = 2\eta\dot{\gamma}_{rz}(r) \quad (2)$$

Hence, from equations (1) and (2)

$$\eta(r) = \frac{\delta P(r)}{\delta r} \frac{h}{4\dot{\gamma}_{rz}(r)} \quad (3)$$

For $N = \frac{1}{2}$, $h \approx 1.3 \times 10^{-2} \text{ cm}$. The value of $\eta(r)$ can be computed from the data in Table 1 for values of r corresponding to $\frac{1}{4}R$, $\frac{1}{2}R$, $\frac{3}{4}R$, R (R being the radius of the hole). The viscosity of ice VI at temperatures between 10 and 16 °C below the melting point and pressures between 1.08 and 1.22 GPa varies from 2.4×10^{13} to $1.4 \times 10^{14} \text{ P}$. The accuracy of our determinations is thought to be no better than $\pm 20\%$, ($\Delta P/P \approx 5\%$, $\Delta h/h \approx 1\%$, $\Delta u/u \approx 10\%$, $\Delta r/r \approx 0$ hence $\Delta \eta/\eta \approx 20\%$; the viscosity found in a repeated experiment differed by 18%) yet the difference is probably significant and the trend is certainly correct indicating an

increase in viscosity with pressure. The activation volume $V^* = RT(\partial \ln \eta)/\partial P$ can be estimated from plotting $\ln \eta$ against P (Fig. 4), it is found to be approximately $V^* \approx 30 \text{ cm}^3 \text{ mol}^{-1}$. If we use this value to extrapolate the viscosity at 2 GPa near the ice VI–VII boundary (and 52 °C below the melting point) we find $\eta \approx 1.7 \times 10^{18} \text{ P}$.

The assumptions used to determine the viscosity probably lead to overestimating rather than underestimating it; the assumption of fully plastic strain regime (compatible with the fact that the volume variation in the deforming regions has experimentally been found close to zero) leads to an upper limit for the shear stress as does the equilibrium condition (1) for a given load. The annealed copper gasket has been found to be just of the right stiffness to allow formation of ice VI without preventing its creeping outwards (platinum was too soft and Inconel too hard). Finally, we assume in equation (2) that viscous flow takes place under the constant plastic shear stress, which is obviously not correct: some stress relaxation must occur.

Finally, in taking $\dot{\gamma} = u/h$, we made the hidden assumption that the velocity profile across the sample is constant up to the vicinity of the anvils where the particles lie. If we make the more reasonable assumption that the profile $u(z)$ is parabolic with $u = 0$ at $z = h/2$ (anvils) and that the particles with the measured velocity u lie at a distance of the anvil surfaces equal to $0.05 h$ (that is $z = 0.45 h$), the strain rate is then $\dot{\gamma} = du/dz \approx 20u/h$, reducing the viscosity by a factor 20.

In these conditions, it is reasonably safe to state that the viscosity of ice VI in most of its domain of existence and within 20 °C of its melting point, is certainly lower than 10^{17} P , and may even be lower than 10^{14} P . Two important consequences follow:

First, if the accretion of Ganymede was homogeneous, subsolidus convection was possible in the ice VI and most of the planet could have remained undifferentiated.

Second, for reasons not taken into account in the current models (such as rapid accretion rate, tidal dissipation and high initial luminosity of Jupiter), melting and differentiation might occur; in that case the liquid mantle could probably freeze again by lithospheric instability or ice diapirism⁵ and transfer heat by solid state convection.

We thank Marc Parmentier for helpful comments. This work was supported by CNRS (Equipe de Recherche Associée no. 886) and by ATP Planétologie INAG no. 37-69. IPG contribution no. 515.

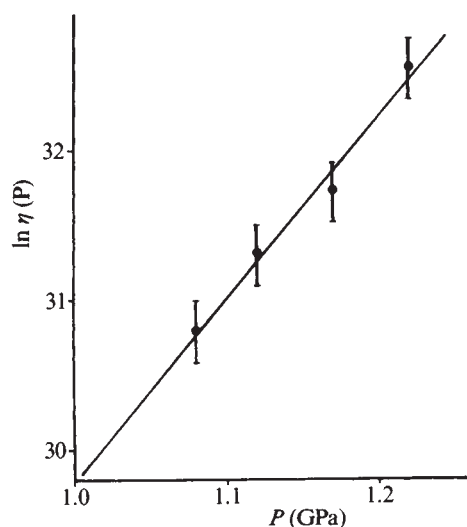


Fig. 4 Plot of the logarithm of the viscosity (η) against pressure (P). The slope of the line gives an activation volume $V^* \approx 30 \text{ cm}^3 \text{ mol}^{-1}$.

Received 5 March; accepted 26 May 1981.

1. Consolmagno, G. J. & Lewis, J. S. *Icarus* **34**, 280–293 (1978).
2. Reynolds, R. T. & Cassen, P. M. *Geophys. Res. Lett.* **6**, 121–124 (1979).
3. Parmentier, E. M. & Head, J. W. *J. geophys. Res.* **84**, 6263–6276 (1979).
4. Squyres, S. W. *Geophys. Res. Lett.* **7**, 593–596 (1980).
5. Parmentier, E. M. & Head, J. W. *Proc. 10th Lunar planet. Sci. Conf.* 2403–2419 (1979).
6. Whatley, L. S. & Van Valkenburg, A. *Advances in High Pressure Research* Vol. 1 (ed. Bradley, R. S.) Ch. 6 (Academic, New York, 1966).
7. Mishima, O. & Endo, S. *J. chem. Phys.* **73**, 2454–2456 (1980).
8. Sung, C. M., Goetze, C. & Mao, H. K. *Rev. scient. Instrum.* **48**, 1386–1391 (1977).
9. Kuriakose, A. K. & Whalley, E. *J. chem. Phys.* **40**, 2025–2031 (1968).
10. Peyronneau, J. & Lacam, A. *Rev. Phys. Appl.* **13**, 107–113 (1978).
11. Pistorius, C. W. *J. Phys. Solids* **26**, 1003–1011 (1965).

Ordering of aluminium and silicon in synthetic faujasites

S. Ramdas, J. M. Thomas* & J. Klinowski

Department of Physical Chemistry, University of Cambridge,
Lensfield Road, Cambridge CB2 1EP, UK

C. A. Fyfe & J. S. Hartman

Guelph-Waterloo Centre for Graduate Work in Chemistry,
University of Guelph, Ontario, Canada N1G 2W1

The distribution of aluminium and silicon atoms on the tetrahedral sites in synthetic faujasites (zeolites X and Y) has been debated for many years. Although the framework structure of the zeolite is known¹, the detailed distribution of Si and Al in the lattice cannot be established by conventional methods, partly because of the very similar scattering powers of silicon and aluminium for X rays and partly because of the difficulties in obtaining crystals of adequate size and perfection. Some inferences have, however, been made on the basis of four indirect approaches: variation of the (cubic) lattice parameter as a function of aluminium content^{2,3}; electrostatic calculations⁴; change of zeolitic acidity on successive extraction of aluminium from the framework⁵⁻⁷; and the kinetics of crystallization⁸. All these approaches have assumed the validity of 'Loewenstein's rule' which forbids Al atoms from occupying neighbouring tetrahedral sites. High-resolution solid-state ²⁹Si NMR (with magic angle spinning) can determine the detailed Si, Al ordering in zeolites⁹⁻¹⁶. NMR coupled with electron diffraction studies has revealed that Loewenstein's rule, generally treated as axiomatic, is broken in Linde A^{10,13,14} and in at least three other zeolites where Si/Al is nominally unity¹⁵. In view of its importance as an archetypal zeolite, particularly commercially, we have re-examined the question of framework ordering in faujasite. We propose new ordering schemes and explain discontinuities in the unit cell constant as a function of composition.

The basis of the solid-state NMR method is the fact that ²⁹Si chemical shifts fall into five distinct ranges, depending on whether a given SiO₄⁴⁻ tetrahedron is linked by oxygen bridges to four, three, two, one or no AlO₄⁵⁻ tetrahedra⁷, these ordering modes are denoted Si(4Al), Si(3Al), Si(2Al), Si(1Al) and Si(0Al), respectively (following ref. 7).

We have carried out a detailed NMR examination of synthetic faujasites with Si/Al ratios in the range 1.18–2.42. Spectra were obtained at 20 °C using a Bruker CXP-100 solid-state high-resolution spectrometer with a home-made magic-angle-spinning attachment. Cross-polarization and proton decoupling were not used. The conical rotors, ~0.5 ml internal volume, were made from Delrin, an acetal resin and spun at a rate of ~3 kHz. The ²⁹Si spectral line widths at half-height were close to 2 p.p.m. All chemical shifts are given from tetramethylsilane, with high field shifts being negative. Time intervals of 1 s between the radio frequency pulses were allowed, and 2,000–10,000 free induction decays were accumulated per sample. Relative intensities of peaks corresponding to the different Si, Al ordering modes have been measured from the spectra and compared with many speculative distributions of Si and Al generated for each crystal composition. Relative populations of the five types of ordering modes were calculated for each such configuration and compared with experiment. We have given special attention to those configurations which do not involve the creation of a net dipole moment, and those associated with the lowest electrostatic energy. The latter has been approximately calculated for a lattice model with fixed point charges on Al sites.

* To whom correspondence should be addressed.

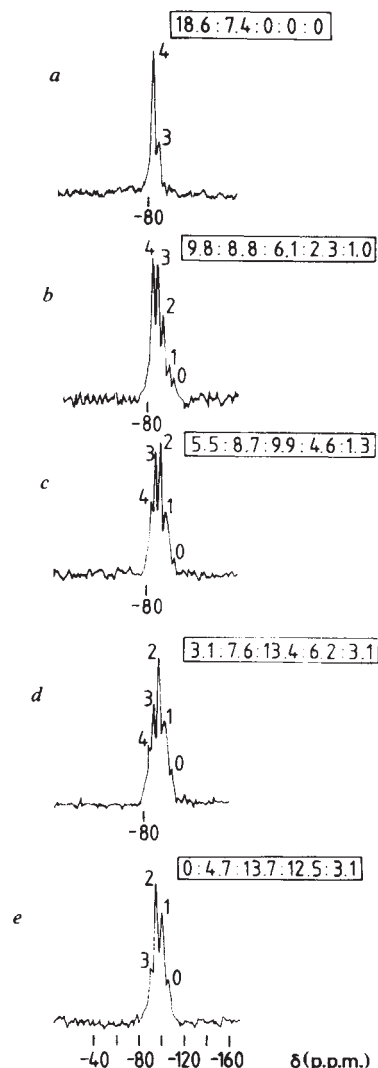


Fig. 1 High-resolution solid-state ²⁹Si NMR spectra of synthetic faujasites. Chemical shifts are given in p.p.m. from tetramethylsilane. Relative populations of the five possible ordering modes, Si(4Al): Si(3Al): Si(2Al): Si(1Al): Si(0Al), estimated from the spectra, are given in the right-hand corner of each box. Peaks corresponding to the respective ordering modes are marked as 4, 3, 2, 1 and 0. The Si/Al ratio for each crystal composition studied, which sometimes slightly deviates from the ideal, is given in the left-hand corner of each box. *a*, Si/Al = 1.19; *b*, Si/Al = 1.35; *c*, Si/Al = 1.67; *d*, Si/Al = 2.00; *e*, Si/Al = 2.45.

Unless restrictions are imposed on the size and symmetry of the unit cell, an infinite number of ordered structures is possible. We have considered a basic unit composed of two sodalite cages linked through a hexagonal prism with a centre of symmetry in the middle. Such basic units are, for heuristic purposes, placed at the points of a face-centred cubic faujasite lattice. Each unit contains 24 pairs of tetrahedral sites. The four-fold multiplicity of the F-lattice allows only certain numbers of Al atoms per unit cell of 192 tetrahedral atoms: 96, 88, 80, 72, 64, 56, and so on, equivalent to Si/Al ratios of 1.00, 1.18, 1.40, 1.67, 2.00, 2.42, and so on, respectively. In faujasite there are two distinct types of six-membered rings in each sodalite cage: four 'type I' rings involved in linkages with neighbouring cages, and four 'type II' (non-linking) rings which face the supercage.

The NMR spectra, given in Fig. 1, show that as the Si/Al ratio approaches unity, the ordering converges towards Si(4Al), that is, towards strict alternation of Si and Al in the lattice. This is in contrast to the situation which prevails in Linde A, where the ordering is Si(3Al) (refs 9, 12 and 13), although the zeolite is also built of sodalite cages. The fact that Loewenstein's rule is broken

in Linde A but obeyed in faujasite will be discussed elsewhere. It would seem reasonable to expect that Al–O–Al linkages would be less likely in less aluminous faujasites, that is, those with higher Si/Al ratios. The perfect ordering in a hypothetical faujasite with Si/Al = 1.00 is given as C1 in Fig. 2. All six-membered rings are of the meta type with respect to both Si and Al. Following Dempsey *et al.*², by now replacing any pair of Al atoms which are related through the centre of symmetry, by Si atoms, thus increasing Si/Al ratio to 1.18, we arrive at configuration C2 with the relative population of the ordering modes of 16:8:0:0:2 which, although the smallest peak cannot be identified, is in good agreement with the intensities in the NMR spectrum (Fig. 1a). Detailed examination revealed that no other speculative configuration can explain NMR results for this composition. The NMR spectra rule out complete disorder.

The transition from Si/Al = 1.18 to Si/Al = 1.40, which is equivalent to altering the Si:Al content per sodalite cage from 13:11 to 14:10, leads to four possible population ratios of ordering modes: 12:8:4:0:4; 10:12:2:0:4; 14:4:6:0:4 (neither shown in Fig. 2) and 8:16:0:0:4 (C3 in Fig. 2), none of which is compatible with the NMR results. However, if we abandon the centre of symmetry requirement and thus the condition of zero dipole moment, we arrive at configurations C4 and C5. C4 agrees very well with experiment (G. Engelhardt, E. Lippmaa and M. Mägi have independently reached the same conclusion) however, C5 (which breaks Loewenstein's rule) does not agree so well with experiment. Approximate calculations of electrostatic energy indicate that either C4 or C5 are acceptable from that point of view. However, the Si (4Al) peak in measured spectra may be higher than Si(3Al) peak, only because the zeolite composition deviates from the ideal Si/Al = 1.40 (see the proportion of these peaks in C2). This means that C5 cannot be unequivocally eliminated.

Closer examination of Fig. 2 reveals that para occupancy of type II rings begins at Si/Al = 1.40 and not at Si/Al = 2.00, as postulated by Dempsey *et al.*². We suggest that this change in mode of occupancy is responsible for the discontinuity in the plot of lattice parameter, occurring at the composition equivalent to 80 Al atoms per unit cell (that is at Si/Al = 1.40). Note that for any composition para occupancy of a type I ring must lead to the violation of Loewenstein's rule if the centre of symmetry is preserved. There is no violation, however, if para occupancy occurs only in type II rings. For Si/Al = 1.40 it is not possible to limit the para occupancy only to type II sites; some para rings of type I have to be admitted if agreement with NMR data is to be achieved. The resultant loss of centrosymmetry gives rise to a net dipole moment. This may, however, be an acceptable consequence, especially if one considers that small adjustments of cationic positions can compensate that dipole moment and hence retain overall non-polarity.

At Si/Al = 1.67 (a 13:9 ratio, that is two Al atoms are eliminated from C5, one of which was involved in an Al–O–Al linkage) configuration C7 is produced with para occupancy in some type I and II rings. C7 is in good agreement with NMR results. However, an alternative configuration, C6, in which para occupancy is restricted to type II rings, also fits the experimental data. Only very accurate measurements, not feasible at present, could distinguish between the two possibilities.

For Si/Al = 2.00 NMR indicates that all type II rings must be of para occupancy (configuration C8). This conclusion agrees with the distribution of Si and Al postulated by Dempsey *et al.*² and evidently accounts for another discontinuity in the cell parameter plot, observed at Si/Al = 2.00. Intensities calculated from C8 agree well with experiment. So do those derived from C9 for Si/Al = 2.42, although good agreement with NMR could be achieved in both cases also by abandoning the $\mu = 0$ condition. In other words, for high Si/Al ratios (from 2.00 upwards), ²⁹Si NMR is a progressively less sensitive tool for the choice of Si, Al ordering. It is, however, invaluable in highly aluminous zeolites, where the correct distribution of tetrahedral atoms can be easily identified.

One further discontinuity in the lattice parameter versus composition plot has been observed², at Si/Al = ~2.80. Although we have no experimental proof, we can speculate, using analogies with more aluminous phases, that the break in the curve is due to the onset of single occupancy of six-membered rings. A configuration in which all six-membered rings are occupied by only one Al atom corresponds to Si/Al = 5.00 and is given in C12. Introduction of partial double occupancy, to reach Si/Al = 3.00 (C10) and Si/Al = 3.80 (C11) involving the necessary existence of meta and para rings, as well as singly-occupied rings, is most likely.

The most reasonable configurations from the many possible are marked with circles in Fig. 2. Unlike any other known and readily applicable technique, ²⁹Si NMR directly probes the immediate environment of the Si atom.

While some of our conclusions may have to be modified when even higher-resolution, higher-sensitivity, solid-state ²⁹Si NMR becomes available, we have already arrived at the progression of

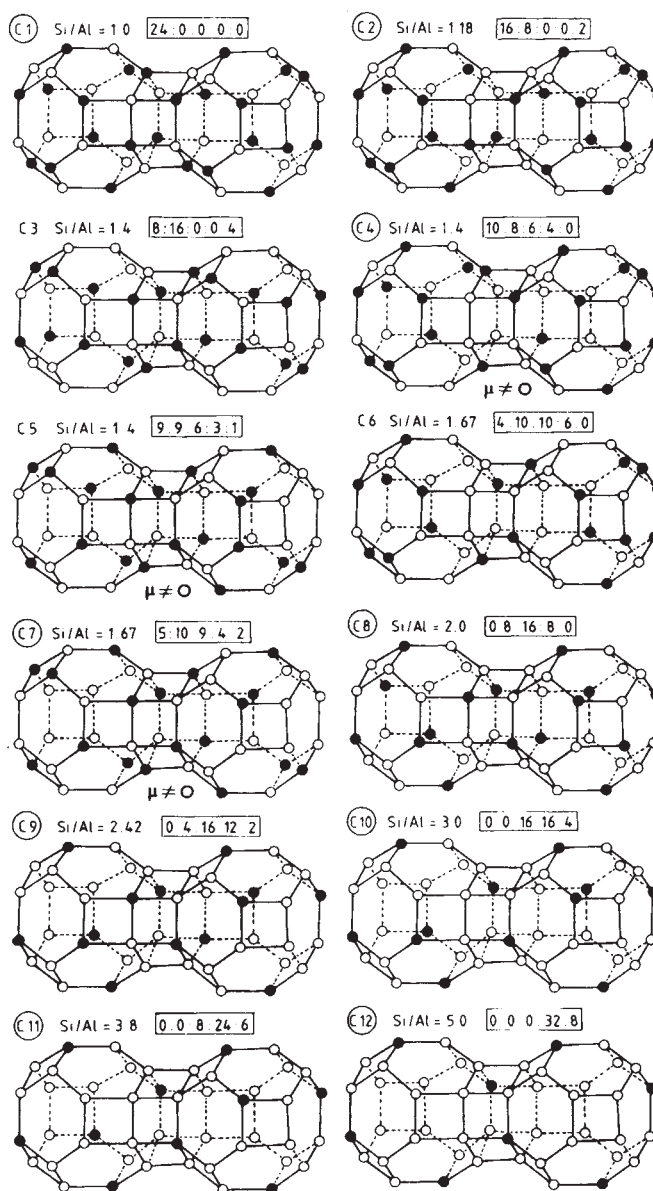


Fig. 2 Possible distributions of Al (●) and Si (○) in the basic structural unit of faujasites with different compositions (given in the left-hand corner of each box). Calculated relative populations of the five possible ordering modes (see Fig. 1) for each diagram are given in the right-hand corner of each box. $\mu \neq 0$ denotes a net dipole moment of the basic structural unit. The most probable configuration for each composition is circled. Configuration C5 violates Loewenstein's rule.

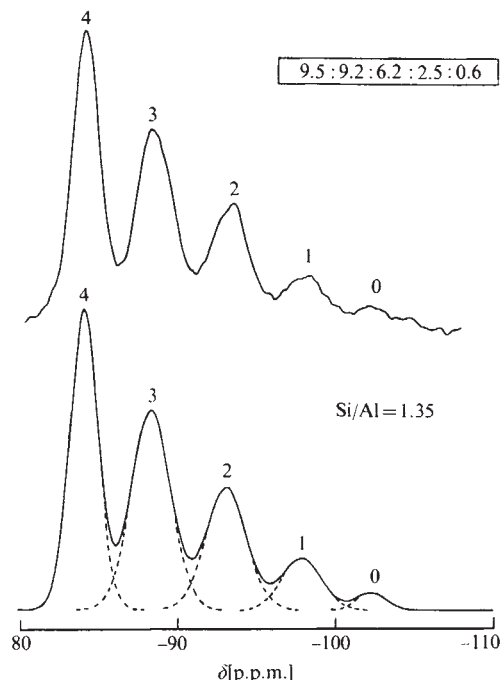


Fig. 3 Observed (at 79.6 MHz) *a*, and simulated *b* high-resolution ^{29}Si NMR spectra for a Si/Al ratio of 1.35.

ordering schemes that give rise to the following unique configurations of Si/Al ratios of 1.0, 1.18, 2.0, 2.42 and 5.0: (1) at Si/Al = 1.0 and 1.18, all meta configurations for six-membered rings of both type I and type II; (2) at Si/Al = 2.0 and 2.42, para occupancies are limited to only the type II rings; and (3) at Si/Al = 5.0, both type I and type II rings have only single occupancies.

In the intermediate configuration, Si/Al = 1.4, para occupancy occurs in some of the type I and type II rings, with consequential generation of a dipole moment. For Si/Al = 1.67, an extension of the ordering schemes applicable for both 1.4 and 2.0 is possible.

Since this work was completed, we have recorded spectra at much higher resolution, and the conclusions drawn from this recent development vindicate our earlier ones. More precise values of the relative populations are now available, and one of the newly recorded spectra, together with its computer-stimulated analogue (based on Gaussian peak shapes) is shown in Fig. 3.

We acknowledge support from the Universities of Cambridge and Guelph, the SRC, BP Research Centre, and Laporte Industries.

Note added in proof: Since this work was submitted we have heard from Engelhardt and Lippmaa that they have obtained essentially the same results and reached the same conclusions¹⁷.

Received 7 May; accepted 15 June 1981.

- Olson, D. H. *J. phys. Chem.* **74**, 2758 (1970).
- Dempsey, E., Köhl, G. H. & Olson, D. H. *J. phys. Chem.* **73**, 387 (1969).
- Breck, D. W. & Flanigen, E. M. in *Molecular Sieves*, 47 (Society of Chemical Industry, London, 1968).
- Dempsey, E. in *Molecular Sieves*, 293 (Society of Chemical Industry, London, 1968).
- Dempsey, E. *J. Catal.* **39**, 155 (1975); **33**, 497 (1974).
- Beaumont, R. & Barthomeuf, D. *J. Catal.* **27**, 45 (1972).
- Olson, D. H. & Dempsey, E. *J. Catal.* **13**, 221 (1969).
- Kacirek, H. & Lechert, H. *J. phys. Chem.* **80**, 1291 (1976).
- Lippmaa, E., Mägi, M., Samoson, A., Engelhardt, G. & Grimmer, A.-R. *J. Am. chem. Soc.* **102**, 4889 (1980).
- Engelhardt, G., Kunath, D., Samoson, A., Tarmak, M. & Mägi, M. *Workshop on Adsorption of Hydrocarbons in Zeolites*, Berlin-Adlershof (1979).
- Lodge, E. A., Bursill, L. A. & Thomas, J. M. *JCS Chem. Commun.* 856 (1980).
- Thomas, J. M., Bursill, L. A., Lodge, E. A., Cheetham, A. K. & Fyfe, C. A. *JCS Chem. Commun.* 276 (1981).
- Bursill, L. A., Lodge, E. A., Thomas, J. M. & Cheetham, A. K. *J. phys. Chem.* (in the press).
- Bursill, L. A., Lodge, E. A. & Thomas, J. M. *Nature* **286**, 111 (1980).
- Klinowski, J., Thomas, J. M., Fyfe, C. A. & Hartman, J. S. (submitted).
- Klinowski, J. *et al.* *JCS Chem. Commun.* (in the press).
- Engelhardt, G. & Lippmaa, E. *Z. anorg. allg. Chem.* (in the press).

Geophysical evidence for non-newtonian gravity

F. D. Stacey & G. J. Tuck

Physics Department, University of Queensland, Brisbane 4067, Australia

Measurements of the variation of gravity with depth in mines and boreholes permit the densities of intervening rock strata to be inferred. In the few cases in which reliable absolute values of density have been independently determined, the calculations can be used to check the value of the newtonian gravitational constant, G . Such large-scale measurements of G are important because the validity of the inverse square law of gravity at short range is being questioned. We have made such a series of measurements and have found four other data sets in the literature that suffice for the estimation of G . We also report here a statistical analysis of 1,100 km² of overlapping sea floor and sea surface gravity data from the Gulf of Mexico (made available by Exxon). All these estimates of G give values that are higher than the conventional, laboratory-determined one. While the possibilities of systematic errors in these data sets preclude a definite conclusion that Newton's law of gravity fails at short range, the strong circumstantial evidence suggests that well controlled large-scale experiments on the inverse square law are urgently required.

There are theoretical reasons for postulating a non-newtonian gravitational effect^{1,2}, and Long's³ claim that, contrary to earlier, less precise measurements⁴, a departure from the inverse square law is observable on a laboratory scale has stimulated novel experiments on gravity⁵⁻⁷. Essentially evidence is sought that, in terms of Newton's law for the attractive force between masses m_1 , m_2 at separation r

$$F = Gm_1m_2/r^2$$

the factor G is not strictly a constant but varies with r . If it exists, the variation is almost certainly small and at astronomical distances it vanishes altogether, but because we have no way of determining the astronomical scale value of G , it is important to extend laboratory type observations to as large a scale as possible.

The earliest estimates of G were geophysical, relying on the attractive effects of mountains of assumed known masses⁸. With the acceptance of the superiority of Cavendish-type laboratory experiments the geophysical attempts became of historical interest only, but now that a scale dependence of G is being sought, geophysical measurements acquire a new significance. They permit the measurement of the gravitational constant over distances of 10–5,000 m, greatly extending the normal laboratory range of 0.05–1 m. Of particular interest are measurements of the vertical gravity gradient in mines, boreholes and at depth within the sea. Airy⁹ first used this method in a coal mine in Durham and it was later pursued more successfully in Czechoslovakia by Von Sterneck¹⁰. The basic idea is that at depth within the Earth the attracting mass is less than that of the whole Earth by the amount of the shell outside the point of measurement. Domzalski¹¹ emphasized instead the potential usefulness of underground gravity surveying as an exploration tool and in the past 30 yr there have been numerous reports of gravity measurements in mines and boreholes, pursuing this idea, but not referring to the possibility of estimating G . The values obtained from those that enable G to be calculated¹²⁻¹⁵ are compared in Table 1 with our own observations in a mine in North-west Queensland, undertaken specifically to determine G , and a new analysis of marine gravity surveys. These seem to be the only reports that provide sufficient information on independent measurements of rock densities, as well as the

Table 1 Values of G from modern geophysical data

Data source	Type of measurement	Gravity data	Depth range (m)	Density data	$(G \pm \sigma)(10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2})$
Ref. 12	Mine	21	96–587	~400	6.795 ± 0.021
Ref. 13	Mine	31	57.3–684.8	47	6.7390 ± 0.0025
		11	57.3–208.5		6.724 ± 0.014
		10	223.1–388.9		6.726 ± 0.012
		10	418.2–684.8		6.746 ± 0.013
		31	57.3–684.8		$*6.7427 \pm 0.0024$
		35	0–684.8		6.7334 ± 0.0037
Ref. 7	Mine	8	0–948	565	6.712 ± 0.037
Ref. 14	Borehole	3	3,712–3,962	16	6.81 ± 0.07
Ref. 15	Mine	7	251–590	53	6.705 ± 0.060
Exxon Exploration Department	Marine surveys	703	113–687		6.797 ± 0.016

*Result obtained with an assumed deep mass anomaly biasing the gravity profile.

variation of gravity with depth, to permit G to be calculated with formal standard deviations that indicate the scatter of the data. The method of analysis follows closely that used in ref. 7. The values in Table 1 should be compared with the conventional laboratory-determined value of $G = (6.672 \pm 0.004) \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$ (ref. 16).

The suspicion that the *in situ* densities of rock strata are systematically underestimated in core and hand samples makes measurements at sea most important. We have been given access to an extensive set of overlapping sea floor and sea surface gravity data from the Gulf of Mexico (Exxon). The surface and sea floor data were obtained independently at different times, but tied to common sea shore reference stations, and both had been free air and Bouguer-corrected to sea level with the assumption that the sea was filled in with sediment. We have tabulated 703 pairs of sea floor values, with corresponding sea surface values and depth, uniformly spaced over an approximately square area of 1,100 km². We are interested in the difference between the surface and sea floor values $g(0)$ and $g(z)$ as a function of water depth, z . The combined effect of the numerical values used in the adjustments is to subtract $0.068223 \text{ mGal ft}^{-1}$ from $[g(z) - g(0)]$. Values of $\Delta g = [g(z) - g(0) - 0.068223 z(\text{ft})]$ are plotted as a function of z in Fig. 1. If the adjustments were perfect, Δg would average to zero for all values of z , so we are interested in the trend apparent in these rather scattered values. The statistical properties are: $n = 703$; $\sum_1^n z = 857782.1 \text{ ft}$; $\sum_1^n \Delta g = -802.5 \text{ mGal}$; $\sum_1^n z^2 = 1203853183 \text{ ft}^2$; $\sum_1^n (\Delta g)^2 = 2744.7754 \text{ mGal}^2$; $\sum_1^n (z \Delta g) = -1266254.728 \text{ mGal ft}$; $\sum_1^n (\Delta g/z) = -554.61526 \text{ mGal ft}^{-1}$; $\sum_1^n (\Delta g/z)^2 = 2303.2955 \text{ mGal}^2 \text{ ft}^{-2}$.

We initially intended to take the linear regression through these data to obtain the departure of dg/dz from the reference

value ($0.068223 \text{ mGal ft}^{-1}$), but a correlation with sea floor topography was noticed. Data from local high points appear systematically above the general trend, although the tendency for data from hollows to give below-average Δg is not as noticeable. The effect may be due to imperfection of the correction for sea floor topography, which was carried out as part of the Bouguer adjustment of the data, but could also arise from a correlation with topography of gravity anomalies from sub-bottom features; in either case it probably gives an appreciable contribution to the gradient of Δg against z . However, the mean value of $\Delta g/z$ is unaffected by topography; positive and negative contributions to $\Delta g/z$ from underlying inhomogeneities average to zero if a large enough area is properly sampled. Using this average, adding the reference gradient and converting to SI units we obtain the result

$$\frac{dg}{dz} = (2.2124 \pm 0.0020) \times 10^{-6} \text{ s}^{-2}$$

Using equation (6) in ref. 7, the theoretical gradient is, with numerical values of the geometrical parameters of the Earth

$$\frac{dg}{dz} = 3.089539 \times 10^{-6} - 12.56616 G \bar{\rho} \text{ s}^{-2}$$

and taking the mean value of the seawater density to be $\bar{\rho} = 1,027 \text{ kg m}^{-3}$, the observed gradient gives the value of G listed as the last item in Table 1. Alternatively, if we attempt to reduce the error resulting from the greater uncertainties of the $\Delta g/z$ values for shallow data by taking the linear regression constrained through the origin, then the $d(\Delta g)/dz = -(0.526 \pm 0.038) \times 10^{-3} \text{ mGal ft}^{-1}$ and the resulting estimate of G is $(6.730 \pm 0.010) \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$.

There are possibilities of systematic errors, not represented by the quoted standard deviations, in all of these estimates. In the case of the marine data we rely on the accuracy of reference baseline ties of the two surveys and there is now no effective way of checking these. In the mine and borehole measurements there is a possibility of a systematic discrepancy between *in situ* and laboratory densities, although this is certainly slight in the case of our own data⁷. We have also considered the possibility of a bias arising from deep anomalous masses. Neither direct measurements⁷ of the free air gradient nor the evidence of a surface gravity survey give evidence of a significant bias. There are sufficient data points in McCulloh's tabulations to provide an internal check and these data are of particular interest because of the small standard deviation that they give. Of the 35 gravity values, we ignored the top 4 which were in the weathered surface layer where densities are unreliable. In Table 1 we have also treated three subsets of 11, 10, 10 of the remaining values. We have also considered the bias of a deep mass anomaly by least square fitting the data with the postulate of an anomalous mass of arbitrary size and depth, without materially affecting the estimate of G .

The evidence of a high value of G is not conclusive but is strongly circumstantial—better controlled experimental data are needed and two new experiments are planned. One is to

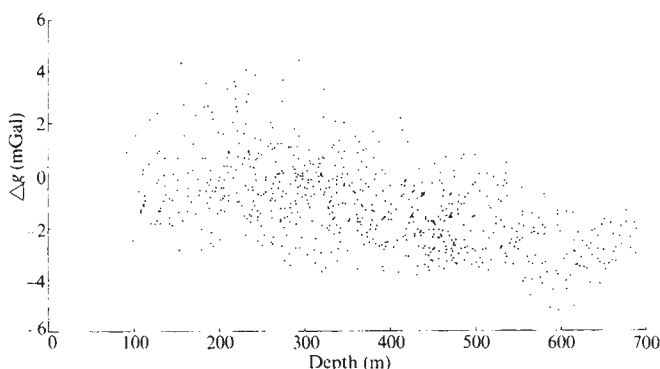


Fig. 1 Gravity differences between overlapping surveys in the Gulf of Mexico. Surface values are subtracted from bottom values after both have been adjusted to correspond to the surface of a sea filled with sediment. If these adjustments were perfect the plotted gravity differences would be zero at all depths, apart from effects of sub-surface features. The scatter of data is due at least partly to sub-surface gravity anomalies.

measure gravity profiles to a depth of about 4 km in topographically and gravitationally flat areas of the ocean¹⁷. The other involves a balance comparison of the weights of 10-kg masses suspended in evacuated tubes at different depths in a lake which will undergo rapid level changes during operation of a hydroelectric pumped storage facility and will offer the possibility of the most precise measurement of G so far.

We thank the Exxon Company for the opportunity to examine gravity records and Michael Alexander and Leroy Brow for helping to make effective use of them.

Received 12 December 1980; accepted 22 May 1981.

1. Fujii, Y. *Nature phys. Sci.* **234**, 5 (1971).
2. Long, D. *Nuovo Cimen.* **55B**, 252 (1980).
3. Long, D. R. *Nature* **260**, 417 (1976).
4. Mackenzie, A. S. *Phys. Rev.* **2**, 321 (1895).
5. Hirakawa, H., Tsubono, K. & Oide, K. *Nature* **283**, 184 (1980).
6. Spero, R., Hoskins, J. K., Newman, R., Pellam, J. & Schultz, J. *Phys. Rev. Lett.* **44**, 1645 (1980).
7. Stacey, F. D., Tuck, G. J., Holding, S. C., Maher, A. R. & Morris, D. *Phys. Rev.* **D23**, 1683 (1981).
8. Bullen, K. E. *The Earth's Density* (Chapman and Hall, London, 1975).
9. Airy, G. B. *Phil. Trans. R. Soc.* **146**, 297, 343 (1856).
10. Von Sterneck, R. in *The Mean Density of the Earth* (ed. Poynting, J. H.) (Charles Griffin, London, 1894).
11. Domzalski, W. *Geophys. Prospect.* **3**, 212 (1955).
12. Whetton, J. T., Myers, J. O. & Smith R. *Geophys. Prospect.* **5**, 20 (1957).
13. McCulloch, T. H. *Geophysics* **30**, 1108 (1965).
14. Hinz, W. J., Bradley, J. W. & Brown, A. R. *J. geophys. Res.* **83**, 5864 (1978).
15. Hussain, A., Walach, G. & Weber, F. *Geophys. Prospect.* (in the press).
16. Cohen, E. R. & Taylor, B. N. *J. phys. Chem. Ref. Data* **2**, 663 (1973).
17. Stacey, F. D. *Geophys. Res. Lett.* **5**, 377 (1978).

Structure of the Earth's inner core

David R. Fearn* & David E. Loper

Geophysical Fluid Dynamics Institute, Florida State University, Tallahassee, Florida 32306, USA

Paul H. Roberts

School of Mathematics, The University, Newcastle upon Tyne NE1 7RU, UK

Jacobs¹ proposed that the Earth's inner core is growing through the freezing of outer-core material as the Earth gradually cools². Recent studies have shown that compositional effects associated with this freezing process can release energy at a rate sufficient to power the geodynamo^{3,4} and may be crucial in determining the dynamic state of the outer core⁵. Here we investigate the effects of composition on the freezing process itself, drawing on metallurgical experience, and speculate on the structure and state of the inner core. We propose that the interface separating the inner and outer core is dendritic and argue that the region in which freezing takes place may extend throughout the entire inner core. Consequently the compositionally driven convective motions which stir the outer core and sustain the geodynamo also occur within the interdendritic spaces of the inner core. The seismic evidence which corroborates this proposal is briefly reviewed.

The material composing the Earth's core has often been treated as being pure iron but it is in fact an "uncertain mixture of all the elements"⁶. The freezing of a mixture can be very different in nature from the freezing of a pure substance⁷ so it is most important that the effects of composition be included in any study of the evolution of the Earth's core. In particular there are two properties which must be modelled. First the temperature (the liquidus temperature T_L) below which material freezes out of a fluid with a certain composition is a function of that composition. Second, in general, the composition of the solid which freezes is different from that of the fluid from which it has

frozen. As a first step towards investigating the effects of composition we shall study the simplest possible system which incorporates these two properties; a binary alloy. We shall assume that as freezing proceeds, one component of the binary mixture goes to form the solid while the other is expelled into the remaining fluid. We shall use ξ to denote the mass fraction of the latter (so by definition $\xi = 0$ in the solid). As already mentioned, the liquidus temperature is a function of composition and we shall use ξ_c to denote that composition ξ for which T_L takes its minimum T_c . In applying the results of studies of binary alloys to the Earth's core, it is very likely⁸ that ξ represents the less dense constituents and the solid which freezes is richer in the denser constituents.

A binary mixture in a single liquid phase has, by Gibbs' phase rule, three independent thermodynamic variables. When a solid phase is present in equilibrium with the adjacent liquid, a constraint is placed on the variables reducing the number of independent variables to two. Usually the constraint is expressed as the liquidus condition $T = T_L(p, \xi)$ where T is the temperature of the liquid and p is the pressure. If

$$T < T_L(p, \xi) \quad (1)$$

the liquid is said to be supercooled and is unstable to the growth of solid, usually in the form of dendrites extending from a boundary. In particular if this condition is caused by a spatial gradient of composition normal to a freezing interface, the liquid is said to be constitutionally supercooled. To determine whether equation (1) is satisfied near the inner-core boundary of the Earth, let us assume the boundary to be spherical and $T = T_L(p, \xi)$ on the boundary. Expanding equation (1) as a Taylor series in powers of $p - p_0$, where p_0 is the pressure at the boundary, and keeping only the linear terms allows equation (1) to be rewritten as

$$(T/L)(\delta - \bar{\mu}\xi d\xi/dp) < dT/dp \quad (2)$$

evaluated at the boundary, where L is the latent heat, δ is the specific-volume change on melting, $\bar{\mu}$ is the energy of mixing⁹, and

$$\partial T_L/\partial p = T\delta/L, \quad \partial T_L/\partial \xi = -T\bar{\mu}\xi/L$$

Because fluid motions are inhibited at the rigid freezing boundary, the fluxes $\xi\dot{m}$ of light material and $L\dot{m}$ of heat, generated by the freezing of material at a mass rate $\dot{m}$ per unit area, must be removed from the interface by diffusion. In addition to the usual flux down the gradient, there is an upward barodiffusive flux of light material, giving^{9,10}

$$\xi\dot{m} = \rho^2 Dg[(\delta/\bar{\mu}) + d\xi/dp] \quad (3)$$

$$L\dot{m} = k\rho g dT/dp \quad (4)$$

where ρ is the fluid density, D the material diffusion coefficient, g the acceleration due to gravity, δ the specific-volume change with composition and k the thermal conductivity. In writing equations (3) and (4) we have assumed a hydrostatic balance to replace radial derivatives by derivatives with respect to pressure. Using equations (3) and (4) to eliminate $d\xi/dp$ and dT/dp from equation (2) yields¹¹

$$\dot{m} > \dot{m}_c = (\delta + \xi\bar{\delta})\rho^2 g D/\bar{\mu}\xi^2 \quad (5)$$

In writing equation (5) a thermal term, which can be shown to be negligible, has been neglected. This is the condition for constitutional supercooling to occur in the liquid outer core immediately above a spherical inner-core boundary.

Using parameter estimates from ref. 11 ($\rho = 1.2 \times 10^4 \text{ kg m}^{-3}$, $g = 4.3 \text{ m s}^{-2}$, $D = 3 \times 10^{-9} \text{ m}^2 \text{ s}^{-1}$, $\bar{\mu} = 4.4 \times 10^7 \text{ J kg}^{-1}$, $\rho\bar{\delta} = 1.1$, $\rho\delta = 0.016$, $\xi = 0.05$), the critical growth rate is $\dot{m}_c = 10^{-10} \text{ kg m}^{-2} \text{ s}^{-1}$. On the other hand, assuming the solid inner core formed 3,000 Myr ago and grew at a constant rate since then, we have $\dot{m} = 5 \times 10^{-8} \text{ kg m}^{-2} \text{ s}^{-1}$. By these estimates, the inner core is growing at a rate roughly 500 times faster than the critical value for the occurrence of constitutional supercooling. Even allowing for a considerable variation in the parameter estimates, it appears that if the inner-core boundary were

* Present address: Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Silver Street, Cambridge CB3 9EW, UK.

spherical, the fluid immediately above it would be constitutionally supercooled. With the above parameter estimates, the supercooling increases with distance from the boundary at a rate of 0.8 K m^{-1} . It is extremely doubtful that such a strongly supercooled fluid could persist for long in the vicinity of a solid crystalline interface. Studies by metallurgists⁷ of the freezing of binary alloys have shown that for growth rates satisfying equation (5), the system reduces the local rate of freezing $\dot{m}$ by increasing the surface area over which freezing occurs, most commonly by forming dendrites extending into the fluid.

The process of freezing and dendritic growth has been described by Copley, *et al.*¹² who studied the evolution of a solution of ammonium chloride in water as it was cooled from below. This solution was studied because it is transparent, allowing the freezing process to be observed directly, and because it has a low entropy of fusion, thus solidifying, as would a metallic solution¹³⁻¹⁵. The cooling causes crystals of ammonium chloride to freeze onto a dendritic interface at the bottom of the solution. The interdendritic liquid is enriched in water and hence is lighter than the overlying solution. This unstable layering drives convective motions despite the stabilizing temperature gradient. The buoyant plumes of lighter fluid erode the dendrites and create isolated chimneys free of solid.

In a quasi-steady state, the cooling system studied by Copley *et al.*¹² may be divided into three regions: a lower region completely solid, an upper region completely liquid and a middle region, the 'mushy zone', in which both solid (dendrites) and liquid are present. The compositionally driven convective motions circulate throughout the upper two regions. The top of the mushy zone is the level at which $T = T_L(p, \xi_o)$ where ξ_o is the mass fraction of the less dense constituent (water) in the well-mixed upper region, while the bottom is the level at which $T = T_L(p, \xi_e)$.

The mushy zone may be thought of as a layer which matches the composition ξ_o of the upper region to the lowest-melting-point composition ξ_e . The thickness of the zone thus depends on the difference $\xi_e - \xi_o$. In the laboratory the dependence of the liquidus temperature on pressure is unimportant and the imposed temperature gradient is sufficiently large that the mushy zone is only a few centimetres deep. In the Earth, pressure effects are important and the temperature gradient is much smaller, causing the mushy zone to be much thicker. To estimate its thickness, the liquidus relation may be rewritten as

$$g\rho\delta dr = -\bar{\mu}\xi d\xi - (L/T) dT \quad (6)$$

Within the mushy zone, $dT/dr < 0$ and a lower bound on its thickness can be found by setting $dT = 0$ in equation (6). Integration yields

$$g_o\rho\delta(r_o^2 - r^2)/r_o > \bar{\mu}(\xi^2 - \xi_o^2) \quad (7)$$

where $g = g_o/r_o$, subscript o denotes conditions at the inner-core boundary (at the top of the mushy zone) and we have assumed $\bar{\mu}$, ρ , δ , r_o and g_o are constant. We may use this inequality to calculate the greatest compositional difference which can be accommodated by a mushy inner core. Setting $r = 0$, $\xi = \xi_e$, $r_o = 1.2 \times 10^6 \text{ m}$ and using the parameter estimates given previously, equation (7) predicts that $\xi_e - \xi_o < 0.016$. This figure is, of course, very uncertain. If we had chosen a larger value of ξ_o , say 0.15, then the allowable difference would have been smaller: $\xi_e - \xi_o < 0.006$. An independent bound on the compositional difference sustained by a mushy inner core can be found using Fig. 5 of ref. 16. With the pressure difference across the outer core known¹⁷ to be $1.93 \times 10^{11} \text{ N m}^{-2}$, the liquidus gradient may be estimated as $(d\xi/dp)_T \leq 10^{-13} \text{ m}^2 \text{ N}^{-1}$. With the pressure difference across the inner core being¹⁷ $3.3 \times 10^{10} \text{ N m}^{-2}$, it follows that $\xi_e - \xi_o \leq 0.003$.

Whether the composition of the core satisfies such an inequality is not known. The point of this calculation is to show that the thickness of the mushy zone is strongly dependent on the difference in composition between the outer core and the corresponding liquid with the lowest melting point. In fact, unless the two compositions are very nearly equal, the mushy

zone will be very thick, possibly extending to the centre of the Earth.

When the idea of compositionally driven convection in the core was first introduced, it was assumed that the resulting flow patterns would be quite similar to those found in thermally driven convection. With the model of a dendritic inner-core boundary, that assumption is not valid. As the core gradually cools, compositionally buoyant fluid is produced within the inner core. This fluid rises upwards into the outer core as thin plumes and is replaced by a general downward motion of heavier fluid. An important feature of this flow pattern, resulting from the fact that diffusion of material is very slow, is that all streamlines must pass through the inner core. Models of fluid flow within the outer core, such as those constructed for kinematic dynamo problems, should have this property.

There may be some seismic evidence which corroborates the view that the inner core is in a partially molten state. Specifically, several authors¹⁸⁻²⁰ have suggested that the attenuation of seismic waves in the inner core may be due to the presence of fluid inclusions. It has also been demonstrated experimentally²¹ that the anelasticity of a simple binary system ($\text{NaCl-H}_2\text{O}$) is much larger if partially molten than if completely solid. Several attempts^{19,20} have been made to demonstrate that thermal diffusive effects in a partially molten inner core can account for the observed anelasticity, but preliminary calculations show that compositional effects are likely to dominate thermal effects; whether these can account for the observed anelasticity remains an open question.

This work was supported by NSF under grant EAR78-00818 and is contribution no. 173 of the Geophysical Fluid Dynamics Institute, Florida State University, Tallahassee, Florida.

Received 9 April; accepted 1 June 1981.

- Jacobs, J. A. *Nature* **172**, 297 (1953).
- Stacey, F. D. *Phys. Earth planet. Inter.* **22**, 89-96 (1980).
- Gubbins, D. J. *Geophys. J.* **43**, 453-464 (1977).
- Loper, D. E. *Geophys. J. R. astr. Soc.*, **54**, 389-404 (1978).
- Fearn, D. R. & Loper, D. E. *Nature* **289**, 393-394 (1981).
- Birch, F. J. *geophys. Res.* **57**, 227-286 (1952).
- Chalmers, B. *Principles of Solidification* (Wiley, New York, 1964).
- Fearn, D. R. & Loper, D. E. in *The Fluid Mechanics of Astrophysics and Geophysics* Vol. 2 (ed. Soward, A. M.) (Gordon and Breach, London, in the press).
- Loper, D. E. & Roberts, P. H. *Geophys. Astrophys. Fluid Dyn.* **9**, 289-321 (1978).
- Loper, D. E. & Roberts, P. H. *Geophys. Astrophys. Fluid Dyn.* **16**, 83-127 (1980).
- Loper, D. E. & Roberts, P. H. *Phys. Earth planet. Inter.* **24**, 302-307 (1981).
- Copley, S. M., Giamei, A. F., Johnson, S. M. & Hornbecker, M. F. *Met. Trans.* **1**, 2193-2204 (1970).
- Jackson, K. A. & Hunt, J. D. *Acta. met.* **13**, 1212-1215 (1965).
- Jackson, K. A., Hunt, J. D., Uhlmann, D. R. & Seward, T. P. *Trans. met. Soc. AIME* **236**, 149-158 (1966).
- Streat, N. & Weinberg, F. *Met. Trans.* **3**, 3181-3184 (1972).
- Usselman, T. M. *Am. J. Sci.* **275**, 291-303 (1975).
- Jacobs, J. A. *The Earth's Core* (Academic, London, 1975).
- Doornbos, D. J. *Geophys. J. R. astr. Soc.* **38**, 397-415 (1974).
- Stiller, H., Franck, S. & Schmit, U. *Phys. Earth planet. Inter.* **22**, 221-225 (1980).
- Cormier, V. F. *Phys. Earth planet. Inter.* **24**, 291-301.
- Spetzler, H. & Anderson, D. L. *J. geophys. Res.* **73**, 6051-6060 (1968).

Iron-manganese banding in Oneida Lake ferromanganese nodules

Willard S. Moore

Department of Geology, University of South Carolina, Columbia, South Carolina 29208, USA

Ferromanganese nodules in the deep-sea and in freshwater lakes usually accrete layers rich in manganese oxides alternating with layers rich in iron oxides¹. The mechanism producing these alternating layers is unknown; indeed, the mechanism producing the nodules themselves is unknown. In Oneida Lake, a large freshwater lake near Syracuse, New York, precipitants from the lake water and the surfaces of nodules at the sediment-water interface are enriched in Mn, whereas nodules buried in lake sediments have surface layers enriched in Fe. It is hypothesized here, using field and laboratory evidence, that reduction and mobilization of Mn from the nodule surface during periods of anoxic sediment cover produce the high Fe layers observed in the nodules.

Oneida Lake nodules have been the subject of several investigations²⁻⁶. Mn phases present are 10 Å manganite (todorokite) and δMnO_2 ; Fe phases are goethite and X-ray amorphous ferric oxides and hydroxides^{4,5}. The nodules exhibit a discontinuous growth pattern characterized by periods of rapid growth interrupted by periods of no growth or erosion⁶. The nodules are most abundant in the 5–10 m deep central area of the lake where ~25% of the bottom may be covered with 1–20 cm diameter diskoidal nodules⁵.

The Mn/Fe ratio of freshly precipitated (<5-yr old) ferromanganese deposits in Oneida Lake is in the range 2.7–3.8 (Table 1). These thin deposits (<1–3 mm) show no Fe-rich layers. The surface Mn/Fe ratios of most Oneida nodules exposed to the Lake water are in this same range (ref. 6 and Table 2). These data suggest that freshly precipitated Oneida ferromanganese deposits including nodule surfaces have a Mn/Fe ratio >2; yet in bulk composition the nodules have Mn/Fe ratios of 1.8 with some layers having ratios as low as 0.07 (ref. 5 and Table 2).

Extensive wind-induced mixing prevents the bottom waters of Oneida Lake from becoming anoxic as many eutrophic lakes do in the summer⁷. Although there is no long-term record of sediment deposition on the central shoals where nodules lie unconformably on a weathering Silurian shale, seasonal events may produce fine-grained sediment cover on these nodules depleting O_2 and leading to the reduction of surficial MnO_2 . In anoxic sediments (those containing <3 $\mu\text{mol O}_2 \text{ kg}^{-1}$) MnO_2 reduction is largely completed before Fe_2O_3 reduction begins^{8,9}. These studies have been confirmed in Oneida Lake sediments¹⁰. Thus nodules buried in sediments should have Mn mobilized from their surfaces in preference to Fe. The Mn may diffuse into the nodule or the adjacent sediment pore water. At site 123 in the lake (see Table 2), such a fine cover was observed during June 1977 but not in August 1977; the sediment cover was absent in August 1978 but present in June 1979. This sediment consisted primarily of organic-rich diatom shells, probably resulting from periods of high productivity in the near-surface waters of the lake⁷. Nodules covered by these anoxic sediments should lose Mn and become enriched in Fe on their surfaces. Removal of the sediment cover by oxidation and winnowing would leave a nodule surface enriched in Fe available to receive more MnO_2 .

The degree of surficial reduction depends directly on the length of time the nodule is covered and the O_2 demand of the sediment. In extreme cases most of the Mn may be removed from a considerable thickness of nodule producing a major Fe-rich band. In other cases Mn accretion may continue relatively uninterrupted for many years to produce a thick Mn-rich band.

There are two sets of data which suggest that these conditions produce the banded structure of the nodules. The first are measurements of low Mn/Fe ratios ranging from 0.07 to 0.31 (Table 2) on the surfaces of nodules recovered from beneath 15–25 cm of sediment in the lake. Below this surface layer, a Mn-rich layer is often present with Mn/Fe ratios of 2–4 (Table 2). Fe-rich layers seem to be forming on these buried nodules. Second, surface depletions of Mn were produced in laboratory experiments by burying nodules in lake sediments for 6 months. In all cases a 1-mm thick layer rich in Fe and depleted in Mn was produced on the nodule surface (Table 2). When the burial

Table 2 Manganese and iron in Oneida Lake nodules

Site no.	Sample description (thickness)	Mn%	Mn/Fe
129	Buried in 15 cm of sediment		
	Hard surface layer (1 mm)	3.44	0.24
	Friable layer below surface (2 mm)	12.5	2.8
129	Buried in 15 cm of sediment		
	Hard surface layer (1 mm)	4.89	0.31
	Friable layer below surface (3 mm)	13.0	3.45
129	Buried in 15 cm of sediment		
	Hard surface layer (1 mm)	0.86	0.07
	Friable layer below surface (2 mm)	7.21	1.16
129	Surficial nodule		
	Friable surface material (1 mm)	17.8	3.03
	Friable material just below surface (3 mm)	16.8	2.63
129	Surficial nodule, appeared highly weathered		
	Very soft surface material	9.27	2.08
	Very hard layer just below surface	1.65	0.10
123	Buried in 25 cm of sediment		
	Flaky surface material (2 mm)	2.79	0.13
	Harder material below flaky surface (2 mm)	4.90	0.25
123	Surficial nodule, appeared weathered		
	Hard layer from surface (1 mm)	4.66	0.18
	Friable material below surface (1 mm)	19.4	1.96
123	Buried in sediment in lab for 6 months		
	Surface layer (1 mm)	9.26	0.27
	Layer below surface (3 mm)	11.01	0.45
123	Buried in sediment in lab for 6 months		
	Surface layer (1 mm)	10.97	0.44
	1st layer below surface (1 mm)	14.86	0.56
	2nd layer below surface (2 mm)	32.33	3.7
123	Buried in sediment in lab for 6 months		
	Surface layer (1 mm)	13.09	0.70
	Layer below surface (2 mm)	32.97	6.1

These sites are 20–30 m south of New York State Barge Canal buoys bearing the same number. Each is a shoal area about 10 m deep with abundant nodules.

experiments were conducted in covered glass jars, a ring of MnO_2 appeared on the glass at the sediment–water interface. In jars of similar sediment but without nodules, no ring of MnO_2 appeared.

The results of these nodule burial experiments are unequivocal. A skin rich in Fe and depleted in Mn is produced by burying a nodule in lake sediments for several months. Apparently MnO_2 on the nodule surface accepts electrons from carbon oxidation in the sediments and is mobilized as Mn^{2+} . In the jar experiments some of this Mn^{2+} diffused to the sediment surface where it was reoxidized to MnO_2 . Because the jars with the sediment but no nodules showed no MnO_2 ring, this Mn must have come from the nodules. No similar Fe mobilization was evident.

Besides explaining the Mn–Fe banding of the nodules, the hypothesis answers an additional question concerning the Oneida nodules, namely their episodic growth. It has been shown⁶ that periods of rapid growth (>1 mm per 100 yr) are separated by growth hiatuses which may last 500–1,000 yr. These growth hiatuses may in fact reflect periods of extensive chemical erosion of the nodules during which Mn is removed selectively and a lag deposit rich in Fe forms on the nodule surface. These hard, Fe-rich layers provide structural integrity for the nodules. The Mn-rich sections are cut easily with a steel blade, but some Fe-rich layers have a hardness >6 enabling the nodules to resist erosion. The partial dissolution of the nodules thus seems to be instrumental in producing a framework strong enough to become an important component of the lake sediments.

I thank Ken Neilson for field assistance and helpful comments, H. Chen for performing the analyses and Dee Hansen for

Table 1 Mn/Fe ratios on different substrates in Oneida Lake

Substrate	Distance above sediment–water interface	Estimated age of deposit	Mn%	Mn/Fe
Plastic bottle	2 m	<2 weeks	22	3.8
Aluminium can	0–4 cm	<2 yr	18	2.7
Glass bottle	0–4 cm	<5 yr	21	3.1
Mussell shell	0–6 cm	<3 yr	19	3.7

typing the manuscript. This work was supported in part by NSF grant OCE 7826846.

Received 6 February; accepted 1 June 1981.

1. Burns, R. G. and Brown, B. A. in *Ferromanganese Deposits on the Ocean Floor* (ed. Horn, D. R.) 51 (National Science Foundation, Washington DC, 1972).
2. Dean, W. E. *Proc. 13th Conf. Great Lakes Res.* 217–226 (1970).
3. Krishnaswami, S. & Moore, W. S. *Nature* **243**, 114–116 (1973).
4. Ghosh, S. K. thesis, Syracuse Univ. (1975).
5. Dean, W. E. & Ghosh, S. K. *J. Res. U.S. geol. Surv.* **6**, 231–240 (1978).
6. Moore, W. S., Dean, W. E., Krishnaswami, S. & Borole, D. V. *Earth planet. Sci. Lett.* **46**, 191–200 (1980).
7. Mills, E. L., Forney, J. L., Clady, M. D. & Schaffner, W. R. in *Lakes of New York State Vol. 2* (ed. Blomfield, J. A.) 367–451 (Academic, New York, 1978).
8. Froelich, P. N. et al. *Geochim. cosmochim. Acta* **43**, 1075–1090 (1979).
9. Klinkhammer, G. P. *Earth planet. Sci. Lett.* **49**, 81–101 (1980).
10. Dean, W. E., Moore, W. S. & Neelson, K. H. *Chem. Geol.* (in the press).

N₂O exchange between a grassland soil and the atmosphere

J. C. Ryden

The Grassland Research Institute, Hurley, Maidenhead, Berkshire SL6 5LR, UK

The recognition that NO, derived from atmospheric N₂O, is an important factor in the destruction of stratospheric ozone¹, and that atmospheric N₂O may reduce radiative heat loss from the Earth's surface², has stimulated interest^{3–6} in the global sources and sinks for N₂O. One possibility is that increasing use of industrially fixed N in agricultural production increases loss of N₂O to the atmosphere. Until recently, however, field data for the evaluation of this concept were not available. Even now, only a few longer-term studies of N₂O loss from representative agricultural land permit estimates of loss over a complete year or a cropping season. In the present study, field measurements of N₂O exchange over 12 months between a grassland soil and the atmosphere indicated that the soil acts as both a source and a sink for atmospheric N₂O, depending on soil conditions and the amount of nitrogenous fertilizer applied.

The measurements of N₂O exchange (as described in Fig. 1) were made in duplicate in a 10 × 10 m subplot within a 0.4-ha plot sown to perennial ryegrass (*Lolium perenne* L.) and receiving 250 kg N ha⁻¹ yr⁻¹ as ammonium nitrate. The field plots formed part of a long-term study of nitrogen fertilizer application to grassland by ICI Ltd at Bracknell, Berkshire, UK. Quadruplicate measurements of N₂O exchange were also made on an unfertilized control strip alongside the 250 kg N ha⁻¹ treatment. The soil at the site is a loam in the Wickham series overlying London clay⁷ (Ochraquaff; US Comprehensive Soil Classification). Some properties of the upper 20 cm of the profile are: pH = 6.0–6.5; organic C = 34.7 mg g⁻¹; total N = 2.3 mg g⁻¹; bulk density = 1.06 g cm⁻³.

Nitrous oxide exchange varied appreciably with time of year and prevailing soil conditions. For the 250 kg N ha⁻¹ treatment (Fig. 1) there was a net loss of N₂O from the soil, weekly mean rates of N₂O loss being as high as 245 ng N m⁻² s⁻¹ (1 ng N m⁻² s⁻¹ = 8.64 × 10⁻⁴ kg N ha⁻¹ day⁻¹) after the third fertilizer application which was made during a period of frequent rain. The highest rates of N₂O loss were observed when the following conditions prevailed simultaneously in the upper 20 cm of the soil profile; water content >20% by weight (air-filled porosity <38% of an undisturbed soil volume), nitrate and ammonium contents >10 µg N per g soil and temperature >10 °C. Additional measurements indicated that the highest rates of N₂O loss were also associated with water content >35% in the upper 2.5 cm of the soil profile.

Low rates of N₂O loss (<12.0 ng N m⁻² s⁻¹) were observed following the first and fourth fertilizer applications. These periods of low N₂O loss coincided with appreciable drying of the soil. Low rates of N₂O loss were also observed during the first 3

weeks of the study even though the soil was at field capacity (water content 28–30%); however, soil temperature was ≤5 °C and nitrate content <1 µg N g⁻¹.

The most striking feature of the data in Fig. 1 is that the 250 kg N ha⁻¹ treatment acted as both a source of and a sink for atmospheric N₂O. Sink activity as high as 11.6 ng N m⁻² s⁻¹ has been measured, although weekly mean values (Fig. 1) never exceeded 5.8 and were generally between 1.2 and 3.5 ng N m⁻² s⁻¹. The prevailing soil conditions during periods of sink activity were moderate to high (>20%) soil water content, very low nitrate content (≤1 µg N g⁻¹) and temperatures above 5–8 °C (Fig. 1). Sink activity was observed before fertilizer application when soil temperature increased above 5 °C. Sink activity resumed 3 weeks after the second fertilizer application and 2 weeks after the third application when soil nitrate content fell to low levels and water content remained

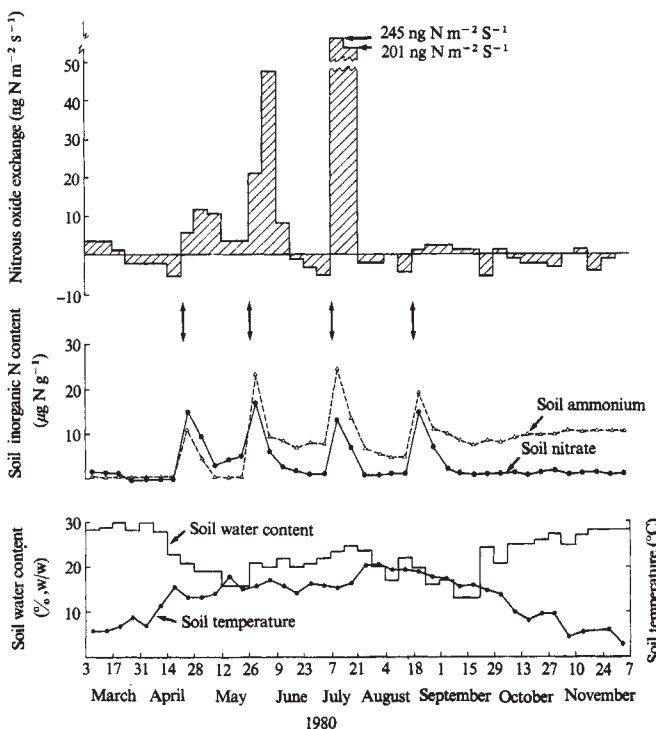


Fig. 1 Nitrous oxide exchange, soil nitrate and ammonium content (0–20 cm), soil moisture content (0–20 cm) and soil temperature (2 cm) in grassland receiving 250 kg N ha⁻¹ yr⁻¹ as ammonium nitrate. Fertilizer was applied in four equal amounts on the dates indicated (↓). The sward was cut and herbage removed immediately before each fertilizer application and on 19 September N₂O exchange was measured using a modified version of the soil cover technique described in ref. 28. The cover had a removable lid, enclosed an air volume of 7.5 dm³ and covered an area of 0.075 m². The air flow (500 cm³ min⁻¹) through the cover was allowed to equilibrate with the soil surface for at least 1 h before starting the N₂O exchange measurement. N₂O exchange was determined for periods of 3 h by comparing the amount of N₂O in air drawn through the cover with that in an equivalent flow of air drawn from the free atmosphere at the same height as the air entry port for the cover. The rate of air flow used to sweep the cover must equal that used to sample the free atmosphere. Accurate matching of flow rates was achieved using stainless steel needle valves and air flow tubes each of which was calibrated to 500 cm³ min⁻¹ ± 1% using a wet gas flowmeter. The air flow control gave a standard error of ±0.005 p.p.m. (v/v) for 10 simultaneous determinations of an atmospheric N₂O concentration of 0.320 p.p.m. The mean atmospheric N₂O concentration during the study period was 0.324 ± 0.013 p.p.m. for 160 determinations. A detection limit of 1.2 ng N m⁻² s⁻¹ for N₂O exchange was obtained. Measurements of N₂O exchange were made on at least 3 days per week between 3 March and 19 September. During periods following fertilizer application or rain, measurements were made on 5 or 6 days per week. After 19 September measurements were made on 2 days per week. To facilitate handling of large numbers of samples, soil nitrate and ammonium content was determined by extraction with 2 M KCl after rapidly drying soils in a forced-air oven at 70 °C. Soil water content was determined gravimetrically on the same samples. Soil nitrate contents determined using this procedure were not significantly different from those determined by extraction of field moist soil. Drying increased ammonium content by 2–4 µg N per g soil.

>20% as a result of persistent rain. Significantly, in a companion treatment receiving 500 kg N ha⁻¹ yr⁻¹ no sink activity was observed—soil nitrate content remained above 2 µg N g⁻¹ throughout the study.

How soil water content affects source or sink activity is illustrated in the 3-week periods following 28 July and 22 September when sink activity ceased as the soil passed through a drying phase. A similar effect was observed in November when the soil passed through periods when its temperature fell below 5 °C. The oscillation between low rates of N₂O loss and sink activity in response to changes in soil temperature continued throughout the winter until the first fertilizer application in mid-March 1981 when significant loss of N₂O resumed.

The unfertilized control strip invariably acted as a sink for N₂O at times when soil water content exceeded 20% (Table 1). Sink activity ceased only when the soil dried to lower water content or when its temperature fell below 5 °C (Table 1; Fig. 1). As with the 250 kg N ha⁻¹ treatment, sink activity was associated with persistently low soil nitrate content, weekly mean and individual values never exceeding 1 µg N g⁻¹ (Table 1).

Measurements made concurrently with those reported in Fig. 1 and Table 1 indicated that denitrification was the major process determining both source and sink activity for N₂O in the present study. Loss of N₂O was always associated with a larger denitrification loss measured using the C₂H₂-inhibition technique⁸ at other areas of the same plot. Rates of N₂O loss from C₂H₂-treated areas were two to seven times greater than those measured in the absence of C₂H₂ (Fig. 1). Production of N₂O during denitrification in soils is favoured by high nitrate content, temperature conducive to high respiratory oxygen demand, and water contents leading to restricted aeration and the development of anaerobic microsites⁹. Maximum N₂O loss was observed (Fig. 1) when such conditions prevailed. During periods of sink activity there was also a small net loss of N (generally <12 ng N m⁻² s⁻¹) from C₂H₂-treated areas. Sink activity was only observed in conditions conducive to microbial reduction of N₂O; namely when the nitrate available for reduction during denitrification was essentially exhausted and when reductive stress was placed on the system^{10,11} with soil water content at or above 20% and temperatures above 5–8 °C.

Some workers have found that N₂O loss may also arise from nitrification, particularly in calcareous soils with high ammonium content¹². Following the application of fertilizer on 7 July 1980, the rates of N₂O loss measured 12–24 h after termination of C₂H₂ treatment were not significantly different (99% confidence level; six replicates) from those measured in areas that had not been treated with C₂H₂ (six replicates). For example, on 11 July, areas treated with C₂H₂ yielded N₂O at

493 ± 135 ng N m⁻² s⁻¹. Other areas that had received C₂H₂ treatment 16 h previously yielded N₂O at 325 ± 24 ng N m⁻² s⁻¹, a rate not significantly different from that (332 ± 58 ng N m⁻² s⁻¹) for areas that had never been treated with C₂H₂. As low concentrations of C₂H₂ also inhibit nitrification, an effect which persists for several days after removal of C₂H₂ (ref. 13), it seems unlikely that nitrification made an appreciable contribution to N₂O loss from this non-calcareous soil even during the periods of high soil ammonium content following fertilizer applications.

The rates of N₂O loss observed in the present study are within a similar range to that reported^{14–21} for arable and grassland systems, although in only one such study¹⁷ has sink activity been observed. Nevertheless, there is direct²² and indirect^{10,23,24} evidence that soil may act as a sink for atmospheric N₂O. Conditions favourable to sink activity^{10,11} may be expected to develop rapidly in moist to wet soils under grass. The dense root mass characteristic of such soils rapidly depletes nitrate in the soil solution and provides potential for appreciable oxygen demand as a result of root respiration and microbial degradation of organic matter. The observation that sink activity rarely exceeded mean weekly rates of 5.8 ng N m⁻² s⁻¹ (Fig. 1; Table 1), irrespective of fertilizer application, suggests that sink activity was limited more by the ease of diffusion of N₂O from the atmosphere to zones of N₂O depletion within the soil, than by the potential for N₂O reduction *per se*.

In the present study, ~1.3% of the fertilizer N applied at a rate of 250 kg N ha⁻¹ yr⁻¹ was lost between 3 March 1980 and 2 March 1981. This proportion is in reasonable agreement with the estimate made by the Council for Agricultural Science and Technology³ for annual N₂O loss from fertilized land but is at least one order of magnitude lower than that implied by other models of N₂O dynamics^{4–6}. The proportion of fertilizer N lost as N₂O in the present study is of the same order as that observed during complete cropping seasons elsewhere^{14–16,18}.

The fact that sink activity for N₂O (Fig. 1; Table 1) was associated with a small net loss of N arising from denitrification (discussed above) suggests that denitrification in soils with low concentrations of inorganic N, despite high total N content, contributes little if any N₂O to the atmosphere. However, additional N₂O may be lost following conversion of grassland to arable production when mineralization of stored fertilizer-derived N may produce elevated concentrations of inorganic N. Nevertheless, even in agricultural systems assumed to be at equilibrium with respect to reserves of soil N, losses of N as N₂O were never greater than 5–7% of the fertilizer N applied even at rates as high as 680 kg N ha⁻¹ yr⁻¹ (ref. 15).

The present and previous studies^{14–20} suggest that forecasts of N₂O losses from industrially fixed N may have been appreciably overestimated. Furthermore, sink activity for N₂O has now been observed in grassland soils (present study), arable soils¹⁷ and lake, estuarine and marine sediments^{25–27}. The indications that N₂O source strength from industrially fixed N may have been overestimated suggest that the sink strength of the pedosphere and hydrosphere for atmospheric N₂O may be more significant than previously assumed.

I thank Drs D. Barraclough and T. R. Owen for cooperation in the use of field facilities, also A. W. Bristow and J. E. McNeill for technical assistance and Dr L. H. P. Jones for helpful comments. The Grassland Research Institute is financed through the Agricultural Research Council.

Received 4 March; accepted 18 May 1981.

- Crutzen, P. J. *Q. J. R. met. Soc.* **96**, 320–325 (1970).
- Wang, W. C. & Sze, N. D. *Nature* **268**, 589–590 (1980).
- Council for Agricultural Science and Technology *Effect of Increased Nitrogen Fixation on Stratospheric Ozone* (Rep. No. 53, Iowa State University, Ames, 1976).
- McElroy, M. B., Wofsy, S. C. & Yung, Y. L. *Phil. Trans. R. Soc. B277*, 159–181 (1977).
- Hahn, J. & Junge, C. Z. *Naturforsch.* **32a**, 190–214 (1977).
- Hahn, J. *Phil. Trans. R. Soc. A290*, 495–504 (1979).
- Jarvis, M. G., Hazelden, J. & Mackney, D. *Soils of Berkshire* (Soil Survey of England and Wales, Harpenden, 1979).
- Ryden, J. C., Lund, L. J., Letey, J. & Focht, D. D. *Soil Sci. Soc. Am. J.* **43**, 110–118 (1979).
- Focht, D. D. & Verstraete, W. *Adv. Microbial Ecol.* **135**–214 (1977).
- Blackmer, A. M. & Bremner, J. M. *Geophys. Res. Lett.* **3**, 739–742 (1976).

Table 1 Weekly mean rates of N₂O exchange between the atmosphere and a grassland soil receiving no fertilizer nitrogen

Week (1980)	Rate of N ₂ O exchange (ng N m ⁻² s ⁻¹)	Soil nitrate content (0–20 cm) (µg N per g soil)	Soil water content (0–20 cm) % (w/w)
18–24 August	–3.1	0.6	20.1
25–31	–1.7	0.8	18.5
1–7 September	–1.2	0.6	19.0
8–14	tr*	tr†	14.1
15–21	tr	0.7	15.0
22–28	–2.9	1.0	22.5
29–5	tr	tr	19.5
6–12 October	–2.3	0.4	20.1
13–19	–1.2	tr	22.2
20–26	–1.2	tr	23.0
27–2	–2.3	0.9	25.2
3–9 November	tr	0.4	23.8
10–16	tr	0.6	24.4
17–23	–4.1	0.9	24.8
24–30	tr	0.5	23.7
1–7 December	tr	0.6	25.2

* <+1.2, >–1.2 ng N m⁻² s⁻¹.

† <+0.1 µg N per g soil.

11. Bremner, J. M. in *Nitrogen in the Environment* Vol. 1 (eds Nielsen, D. R. & MacDonald, J. G.) 474–491 (Academic, New York, 1978).
12. Bremner, J. M. & Blackmer, A. M. *Science* **199**, 295–297 (1978).
13. Walter, H. M., Keeney, D. R. & Fillery, I. R. *Soil Sci. Soc. Am. J.* **43**, 195–196 (1979).
14. Hutchinson, G. L. & Mosier, A. R. *Science* **205**, 1125–1127 (1979).
15. Ryden, J. C. & Lund, L. J. *J. envir. Qual.* **9**, 387–393 (1980).
16. McKenney, D. J., Shuttleworth, K. F. & Findlay, W. I. *Can. J. Soil Sci.* **60**, 429–438 (1980).
17. Bremner, J. M., Robbins, S. G. & Blackmer, A. M. *Geophys. Res. Lett.* **7**, 641–644 (1980).
18. Burford, J. R., Dowdell, R. J. & Cress, R. J. *Sci. Fd Agric.* **32**, 219–223 (1981).
19. Conrad, R. & Seiler, W. *Atmos. Envir.* **14**, 555–558 (1980).
20. Breitenbeck, G. A., Blackmer, A. M. & Bremner, J. M. *Geophys. Res. Lett.* **7**, 85–88 (1980).
21. Denmead, O. T., Freney, J. R. & Simpson, J. R. *Soil Sci. Soc. Am. J.* **43**, 726–728 (1979).
22. Cicerone, R. J., Shetter, J. D., Stedman, D. M., Kelly, T. J. & Liu, S. C. *J. geophys. Res.* **83**, 3042–3050 (1978).
23. Letey, J., Hadas, A., Valoras, N. & Focht, D. D. *J. envir. Qual.* **9**, 232–235 (1980).
24. Delwiche, C. C., Bissell, S. & Virginia, R. in *Nitrogen in the Environment* Vol. 1 (eds Nielsen, D. R. & MacDonald, J. G.) 459–476 (Academic, New York, 1978).
25. Chan, Y. K. & Knowles, R. *Appl. envir. Microbiol.* **37**, 1067–1072 (1979).
26. Elkins, J. W., Wofsy, S. C., McElroy, M. B., Kolb, C. E. & Kaplan, W. A. *Nature* **275**, 602–606 (1978).
27. Cohen, Y. *Nature* **272**, 235–237 (1978).
28. Ryden, J. C., Lund, L. J. & Focht, D. D. *Soil Sci. Soc. Am. J.* **42**, 731–738.

Pseudogenes as a paradigm of neutral evolution

Wen-Hsiung Li, Takashi Gojobori & Masatoshi Nei

Center for Demographic and Population Genetics,
The University of Texas at Houston, Houston, Texas 77025, USA

On the neutral mutation hypothesis^{1–3}, the rate of nucleotide substitution is expected to be higher for functionally less important genes or parts of genes than for functionally more important genes, as the latter would be subject to stronger purifying (negative) selection^{2,4}. On the other hand, selectionists believe that most nucleotide substitutions are caused by positive darwinian selection^{5,6}, in which case the rate of nucleotide substitution in functionally unimportant genes or parts of genes^{2,7} is expected to be relatively lower because the mutations in these regions of DNA would not produce any significant selective advantages. Kimura⁸ and Jukes⁹ have argued that the higher substitution rate observed at the third positions of codons than at the first two positions supports the neutral mutation hypothesis, as most third-position substitutions are synonymous and do not change the amino acids encoded, although others^{5,10} have discussed the possibility that third-position substitutions are subject to positive darwinian selection. Recently, Kimura¹¹ noted that the mouse globin pseudogene, $\psi\alpha 3$, evolved faster than the normal mouse $\alpha 1$ gene, although he did not compute the substitution rate. Here, we present a method of computing the rate of nucleotide substitution for pseudogenes, and report that the three recently discovered pseudogenes show an extremely high rate of nucleotide substitution. As these pseudogenes apparently have no function, this finding strongly supports the neutral mutation hypothesis.

A pseudogene is a DNA segment with high homology with a functional gene but containing nucleotide changes such as frameshift and nonsense mutations that prevent its expression. (Some authors¹² have argued possible functions of pseudogenes, but their arguments are not substantiated.) Pseudogenes seem to have been produced by the nonfunctionalization of duplicate genes. A complete nucleotide sequence is now available for three pseudogenes (mouse $\psi\alpha 3$, human $\psi\alpha 1$ and rabbit $\psi\beta 2$ in the globin gene families)^{13–15}. Figure 1 shows a probable evolutionary scheme for mouse pseudogene $\psi\alpha 3$ ($M\psi\alpha 3$), mouse functional gene $\alpha 1$ ($M\alpha 1$)¹⁶ and rabbit functional gene α ($R\alpha$)¹⁷ in which O denotes the point of duplication leading to $M\psi\alpha 3$ and $M\alpha 1$.

Let d_{ABi} , d_{ACi} and d_{BCi} be the numbers of nucleotide substitutions per site at the i th position of codons ($i = 1, 2$ or 3) between $M\psi\alpha 3$ and $M\alpha 1$, between $M\psi\alpha 3$ and $R\alpha$, and between $M\alpha 1$ and $R\alpha$, respectively, and let l_i , m_i and n_i be the numbers of nucleotide substitutions per site between O and $M\psi\alpha 3$, between O and $M\alpha 1$ and between O and $R\alpha$, respectively. We then have $d_{ABi} = l_i + m_i$, $d_{ACi} = l_i + n_i$, $d_{BCi} = m_i + n_i$. Therefore, l_i , m_i and n_i can be estimated by

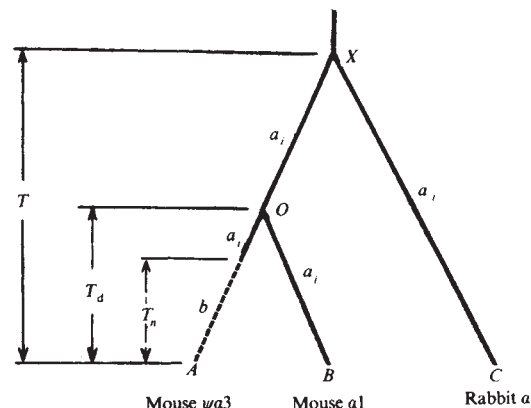


Fig. 1 Plausible phylogenetic tree for mouse $\psi\alpha 3$, mouse $\alpha 1$ and rabbit α . T denotes the divergence time between mouse and rabbit, T_d the time since duplication of mouse $\psi\alpha 3$ and $\alpha 1$, and T_n the time since nonfunctionalization of mouse $\psi\alpha 3$. a_i denotes the rate of nucleotide substitution per site per year at the i th position of codons in the normal globin genes and b the rate of substitution for mouse pseudogene $\psi\alpha 3$. The proportion of nucleotide differences is 51/354 between mouse $\psi\alpha 3$ and $\alpha 1$, 84/423 between mouse $\alpha 1$ and rabbit α , and 91/354 between mouse $\psi\alpha 3$ and rabbit α (see Table 1).

$$l_i = (d_{ABi} + d_{ACi} - d_{BCi})/2 \quad (1a)$$

$$m_i = (d_{ABi} - d_{ACi} + d_{BCi})/2 \quad (1b)$$

$$n_i = (-d_{ABi} + d_{ACi} + d_{BCi})/2 \quad (1c)$$

Pseudogenes often have deletions and insertions. The nucleotides involved in these changes should be eliminated from data analysis, because we are interested only in nucleotide substitution. This can be done by aligning a pseudogene with its homologous functional gene for the coding regions. In our study we followed the alignments given by the original authors^{13–15}. In mouse $\psi\alpha 3$, however, we excluded the 30 nucleotides that have been aligned with the nucleotides starting from positions 91 to 120 of the functional mouse $\alpha 1$. In this region there were 21 mismatches, including one sequence of 8 mismatches and another of 7 mismatches. This pattern was very different from that of other parts of the gene and probably occurred through insertion rather than nucleotide substitution. At any rate, after these alignments, we computed the proportion of different nucleotides between homologous genes (Table 1), and from this proportion (p) the total number of nucleotide substitutions per site (d_{ABi} , d_{ACi} or d_{BCi}) was estimated by $d = -(3/4) \ln [1 - (4/3)p]$ (ref. 18). For computing d we also

Table 1 Proportions of nucleotide differences between genes for the three positions of codons

Genes compared	First position	Second position	Third position
$M\psi\alpha 3$ with: $M\alpha 1$	12/118	17/117	22/119
$H\alpha$	19/118	23/117	47/119
$R\alpha$	21/118	22/117	48/119
$H\psi\alpha 1$ with: $H\alpha$	30/133	26/132	42.5/133
$M\alpha 1$	34/133	31/132	49/133
$R\alpha$	39/133	32/132	51/133
$M\alpha 1$ with: $H\alpha$	13/141	14/141	54.5/141
$R\alpha$	16/141	15/141	53/141
$H\alpha$ with: $R\alpha$	18/141	14/141	31/141
$R\psi\beta 2$ with: $R\beta 1$	26/146	19/145	34/145
$H\beta$	31/146	24/145	42/145
$M\beta$	34/146	31/145	43/145
$R\beta 1$ with: $H\beta$	9/146	6.5/146	32/146
$M\beta$	21.5/146	16/146	46/146

$M\psi\alpha 3$ = mouse $\alpha 3$ pseudogene; $M\alpha 1$ = mouse $\alpha 1$; $H\alpha$ = average for human $\alpha 1$ and $\alpha 2$; $R\alpha$ = rabbit α ; $H\psi\alpha 1$ = human $\alpha 1$ pseudogene; $R\psi\beta 2$ = rabbit $\beta 2$ pseudogene; $R\beta 1$ = average for rabbit $\beta 1$ allele 1 and allele 2; $H\beta$ = human β , $M\beta$ = mouse β major. The 30 bases near the middle of $M\psi\alpha 3$ are excluded from the comparison (see text).

Table 2 Numbers of nucleotide substitutions per site at the first, second and third positions of codons between *O* and *A*, between *O* and *B* and between *O* and *C*, where *O* is the point of duplication leading to sequences *A* and *B* in Fig. 1

Pseudogene (Sequence A)	Sequence B	Sequence C	Between <i>O</i> and <i>A</i>			Between <i>O</i> and <i>B</i>			Between <i>O</i> and <i>C</i>		
			1	2	3	1	2	3	1	2	3
Mouse $\psi\alpha 3$	Mouse $\alpha 1$	Human α , rabbit α	0.095	0.137	0.125	0.014	0.025	0.088	0.097	0.086	0.446
Human $\psi\alpha 1$	Human α	Mouse $\alpha 1$, rabbit α	0.246	0.205	0.268	0.022	0.024	0.148	0.097	0.083	0.254
Rabbit $\psi\beta 2$	Rabbit $\beta 1$	Human β , mouse β^{maj}	0.184	0.140	0.159	0.020	0.004	0.122	0.086	0.080	0.216

used Kimura's formula¹¹, but the values were not much different.

In Fig. 1 we used rabbit α as a third gene, but human $\alpha 1$ or $\alpha 2$ can also be used for this purpose, as their nucleotide sequences are known^{14,19}. We have therefore computed the d values for each of these genes and used the averages of the values to compute l_i , m_i and n_i (Table 2) and the other quantities to be given later. Table 2 also includes the results for human pseudogene $\psi\alpha 1$ and rabbit pseudogene $\psi\beta 2$. It is seen that in the evolution of functional genes (between *O* and *B* and between *O* and *C*), the third position of codons changes several times faster than the first two positions, which evolve at almost the same rate, but in the line leading to a pseudogene (between *O* and *A*) the rates of change for the three positions are about the same and are higher than those between *O* and *B*. Clearly, pseudogenes evolve faster than functional genes.

The above computation does not give the substitution rate for pseudogenes. To compute this rate, we must know the time since nonfunctionalization of a pseudogene. Let us again use the three genes in Fig. 1 as an example. In this figure the time (T) since divergence between $M\alpha 1$ and $R\alpha$ is known to be about 80 Myr, but the time (T_d) since duplication of $M\psi\alpha 3$ and $M\alpha 1$ and the time (T_n) since nonfunctionalization of $M\psi\alpha 3$ must be estimated. We estimate these times and the rates of nucleotide substitution for the functional genes and pseudogenes simultaneously. Let a_1 , a_2 and a_3 be the rates of nucleotide substitution per site per year at the first, second and third positions of codons in the functional genes, respectively. Once a gene becomes nonfunctional, the rate of nucleotide substitution is expected to be the same for all of the three positions, and we denote the rate by b . From Fig. 1 we obtain

$$d_{ABi} = 2a_i T_d + (b - a_i) T_n \quad (2)$$

$$d_{ACi} = 2a_i T + (b - a_i) T_n \quad (3)$$

$$d_{BCi} = 2a_i T \quad (4)$$

As we know T , a_i can be estimated by $d_{BCi}/(2T)$. We also note

$$y_i \equiv d_{ACi} - d_{BCi} = bT_n - a_i T_n \quad (5)$$

Therefore, from equations (2) and (5), T_d can be estimated by $(\sum d_{ABi} - \sum y_i)/(2\sum a_i)$, where $\sum$ stands for the summation over i .

To estimate T_n and b , we can use equation (5) and apply the standard least-squares method, as there are two unknowns and three equations. However, note that essentially the same results are obtained by the following simple formulae

$$T_n = (y_{12} - y_3)/(a_3 - a_{12}) \quad (6)$$

$$b = (a_3 y_{12} - a_{12} y_3)/(y_{12} - y_3) \quad (7)$$

where $y_{12} = (y_1 + y_2)/2$ and $a_{12} = (a_1 + a_2)/2$. Expressing equations (6) and (7) in terms of l_i , m_i and n_i , we have obtained approximate formulae for the standard errors of T_n and b (not

shown). In practice, T_n obtained by equation (6) may be larger than T_d . In this case, we set $T_d = T_n$ because by definition $T_n \leq T_d$. When $T_n = T_d$, equation (5) gives $b = (\sum a_i T_d + \sum y_i)/T_d$.

Table 3 shows the results for a_i , T_d , T_n and b . Interestingly, each of a_1 , a_2 and a_3 is nearly the same for the three groups of genes studied. However, a_3 is about four times larger than a_1 and a_2 . These estimates are similar to those obtained by Kimura⁸ and Jukes⁹ and support these authors' conclusion that the rate of synonymous substitution is higher than the rate of non-synonymous substitution. The b values in Table 2 indicate that the rate of nucleotide substitution in pseudogenes is even higher than the rate at the third positions of codons in the functional genes. The average rate of 4.6×10^{-9} is one of the highest rates of nucleotide substitutions so far estimated. Only two other estimates are comparable with this value. One is the rate (7.0×10^{-9}) for the synonymous substitution in the C-peptide region of the preproinsulin genes²⁰, and the other is that (6.2×10^{-9}) estimated from amino acid sequence data for the rapidly evolving residues of fibrinopeptides²¹. These peptides are believed to have no biological function except for holding other polypeptides that will later form a protein. Our finding clearly indicates that functionally less important genes evolve faster than functionally more important genes, and thus supports the neutral mutation hypothesis. Furthermore, comparison of a_3 and b suggests that the third-position substitutions in the globin genes are subject to purifying selection.

Our estimate (T_d) of the time since gene duplication is 27 Myr for mouse $\psi\alpha 3$ and $\alpha 1$, 49 Myr for human $\psi\alpha 1$ and α , and 44 Myr for rabbit $\psi\beta 2$ and $\beta 1$. These estimates are somewhat smaller than those (30, 60, 55 Myr, respectively) obtained by Maniatis and his associates^{14,15}, mainly because these authors assumed that the substitution rate for the pseudogenes is the same as that for the synonymous changes in functional globin genes. This assumption seems to be incorrect, as the former rate is about twice as great as the latter rate in our study. If our estimates of the times of gene duplication are reliable, the mouse $\psi\alpha 3$ diverged from the $\alpha 1$ gene about 27 Myr ago and became a pseudogene about 4 Myr later. On the other hand, the human $\psi\alpha 1$ was duplicated from the α gene about 49 Myr ago and became nonfunctional about 4 Myr later. The rabbit $\psi\beta 2$ gene became nonfunctional almost immediately after it was duplicated from the $\beta 1$ gene about 44 Myr ago.

We have studied the evolutionary changes of pseudogenes as a paradigm of neutral evolution because in these genes the fate of new mutations in a population is determined almost entirely by genetic drift under the neutral mutation hypothesis. In practice, most genes in the genome would have some biological function and thus be subject to a varying degree of purifying selection. Even in these genes, however, there may be a large

Table 3 Times since gene duplication (T_d), times since nonfunctionalization (T_n) and rates of nucleotide substitution per site per year (b) for pseudogenes mouse $\psi\alpha 3$, human $\psi\alpha 1$ and rabbit $\psi\beta 2$

Pseudogene (Sequence A)	Sequence B	Sequence C	a_1 ($\times 10^{-9}$)	a_2 ($\times 10^{-9}$)	a_3 ($\times 10^{-9}$)	T_d (Myr)	T_n (Myr)	b ($\times 10^{-9}$)
Mouse $\psi\alpha 3$	Mouse $\alpha 1$	Human α , rabbit α	0.69 ± 0.26	0.69 ± 0.26	3.32 ± 0.73	27 ± 6	23 ± 19	5.0 ± 3.2
Human $\psi\alpha 1$	Human α	Mouse $\alpha 1$, rabbit α	0.74 ± 0.27	0.67 ± 0.26	2.51 ± 0.61	49 ± 8	45 ± 37	5.1 ± 3.3
Rabbit $\psi\beta 2$	Rabbit $\beta 1$	Human β , mouse β^{maj}	0.71 ± 0.27	0.51 ± 0.22	2.09 ± 0.51	44 ± 8	44*	3.6*
Average			0.71	0.62	2.64			4.6

a_1 , a_2 and a_3 refer to the rates of nucleotide substitutions for the first, second and third positions of codons in the functional genes, respectively.

* We were unable to compute a proper standard error for this estimate, because T_n was assumed to be equal to T_d (see text).

number of mutant forms (nucleotide sequences) that are equally functional and fit in adaptation. All these mutant forms will be a source of neutral evolution.

This work was supported by grants from the NIH and NSF.

Note added in proof: After submission of this paper, we learned that Miyata and Yasunaga²² estimated the rate of nucleotide substitution for mouse $\psi\alpha 3$. Their estimate is higher than ours, because they did not exclude the middle part of the sequence.

Received 17 February; accepted 5 May 1981.

- Kimura, M. *Nature* **217**, 624–626 (1968).
- King, J. L. & Jukes, T. H. *Science* **164**, 788–798 (1969).
- Kimura, M. & Ohta, T. *Proc. natn. Acad. Sci. U.S.A.* **71**, 2848–2852 (1974).
- Dickerson, R. E. *J. molec. Evol.* **1**, 26–45 (1971).
- Clarke, B. *Science* **168**, 1009–1011 (1970).
- Milkman, R. *Trends biochem. Sci.* **1**, N152–N154 (1976).
- Jukes, T. H. & King, J. L. *Nature* **231**, 114–115 (1971).
- Kimura, M. *Nature* **267**, 275–276 (1977).
- Jukes, T. H. *J. molec. Evol.* **11**, 207–209 (1978).
- Richmond, R. C. *Nature* **225**, 1025–1028 (1970).
- Kimura, M. *J. molec. Evol.* **16**, 111–120 (1980).
- Proudfoot, N. J. *Nature* **286**, 840–841 (1980).
- Nishioka, Y., Leder, A. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2806–2809 (1980).
- Proudfoot, N. J. & Maniatis, T. *Cell* **21**, 537–544 (1980).
- Lacy, E. & Maniatis, T. *Cell* **21**, 545–553 (1980).
- Nishioka, Y. & Leder, P. *Cell* **18**, 875–882 (1979).
- Heindell, H. C. *et al. Cell* **15**, 43–54 (1978).
- Jukes, T. H. & Cantor, C. H. *Mammalian Protein Metabolism* (ed. Munro, H. N.) 21–123 (Academic, New York, 1969).
- Michelson, A. M. & Orkin, S. H. *Cell* **22**, 371–377 (1980).
- Perler, F. *et al. Cello* **20**, 555–566 (1980).
- Kafatos, F. C. *et al. Proc. natn. Acad. Sci. U.S.A.* **74**, 5618–5622 (1977).
- Miyata, T. & Yasunaga, T. *Proc. natn. Acad. Sci. U.S.A.* **78**, 450–453 (1981).

Gait and the energetics of locomotion in horses

Donald F. Hoyt* & C. Richard Taylor

Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138, USA

It seems reasonable that quadrupeds should change gait from a walk to a trot to a gallop in such a way as to minimize their energy consumption, as human beings are known¹ to change from a walk to a run at a particular speed (2.4 m s^{-1}) below which walking requires less energy than running and above which the opposite is true. Thus by changing gait, human beings keep the energy cost of locomotion to a minimum as their speed increases. One reason this relation holds is that in humans, metabolic rate increases curvilinearly with walking speed¹. If metabolism were a curvilinear function of speed within each of the gaits used by quadrupeds, it would support the hypothesis that they also change gait to minimize energetic cost. There is an old controversy about whether metabolic rate increases linearly or curvilinearly in running humans^{1,2} but all previous reports have suggested that metabolic rate increases linearly with speed in quadrupeds. Extended gaits were an important experimental tool in the study of human gait changes; thus we have trained three small horses (110–170 kg) to walk, trot and gallop on a motorized treadmill, and to extend their gaits on command. We report here that, using measurements of rates of oxygen consumption as an indicator of rates of energy consumption, we have confirmed that the natural gait at any speed indeed entails the smallest possible energy expenditure.

Rate of oxygen consumption increased curvilinearly with speed for walking and trotting (Fig. 1). We were unable to obtain sufficiently high galloping speeds to evaluate whether rate of oxygen consumption also increased curvilinearly during a gallop. Transitions between gaits normally occurred at the speeds where the curves intersected and oxygen consumption was the same for the two gaits. When the gaits were extended beyond their normal range of speeds, oxygen consumption was higher in the extended gait than in that which the animal would normally be using. For example, horse B normally walked at a speed of 1.25 m s^{-1} . When trotting at 1.25 m s^{-1} , the rate of oxygen consumption was 1.5 times that measured during a walk.

* To whom correspondence should be addressed at: Department of Biological Sciences, California State Polytechnic University, Pomona, California 91768, USA.

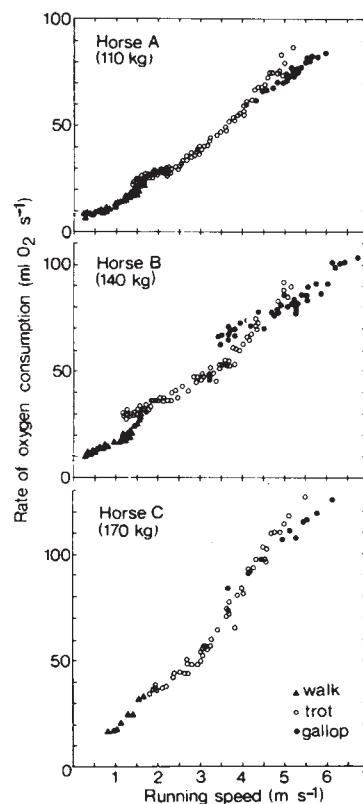


Fig. 1 Steady-state rate of oxygen consumption as a function of speed for a walk, a trot and a gallop in three small horses. The horses normally changed from walk to trot, or trot to gallop at the point where rate of oxygen consumption was the same in the two gaits. We were able to train the animals to extend their gaits to speeds where they would have normally used a different gait. The data from the extended gaits accentuated the curvilinear nature of the relationship between rate of oxygen consumption and speed within a gait. Linear regression provided a good fit of the data if we omitted measurements made when animals used extended gaits. The least-squares regressions and coefficients of determination (r^2) are: horse A, $\dot{V}_{O_2} = -1.22 + 14.4 \cdot v$, $r^2 = 0.99$; B, $\dot{V}_{O_2} = 2.68 + 15.0 \cdot v$, $r^2 = 0.98$; C, $\dot{V}_{O_2} = -8.61 + 22.2 \cdot v$, $r^2 = 0.98$ where $\dot{V}_{O_2}$ is $\text{ml O}_2 \text{ s}^{-1}$ and v is velocity in m s^{-1} . We considered that a steady-state oxygen consumption was achieved when we recorded a constant rate of oxygen consumption for at least 5 min. At the speeds used here, steady-state was never reached during the first 6 min of measurement, thus all measurements at a given speed were made for at least 11 min. During these measurements, air temperature averaged 21°C and air speed was approximately matched to tread speed. The animals wore loose-fitting masks which allowed capture of all expired gases. Rates of gas flow through the mask were varied from 150 l min^{-1} to $1,300 \text{ l min}^{-1}$ during different runs to produce excurrent oxygen concentrations between 20.6 and 20.1%. During the runs the laboratory was ventilated by a large fan and ambient oxygen concentrations remained constant at $\sim 20.9\%$. Previous studies of horses, including pony B of ref. 4, found no change in blood lactate levels when animals ran on a level treadmill at speeds up to 10 m s^{-1} ; therefore we conclude that there is no anaerobic contribution to energy consumption at speeds $< 7 \text{ m s}^{-1}$.

There was a speed for each gait where the amount of oxygen used to move a given distance (rate of oxygen consumption divided by speed) reached a minimum value (Fig. 2). This speed represents one that is energetically optimal for each gait. The minimum values were about the same for a walk, a trot and a gallop. Therefore, the amount of energy consumed by a horse to move a given distance is the same at these optimal speeds.

To determine whether animals normally moved at energetically optimal speeds, we measured speeds and gaits of a horse as it moved freely over a marked grid. We found that the horse selected speeds within each gait around the energetically optimal speed (Fig. 2). Pennycuik³ has observed that migrating African animals use only a restricted range of speeds within each gait. This suggests that these animals might also be using an energetically optimal speed for each gait.

Linear regression of the data provides a reasonable fit to the relationship between oxygen consumption and speed if one disregards the data obtained when the animals had abnormally extended gaits ($r^2 = 0.98$). In fact the curvilinear relationship

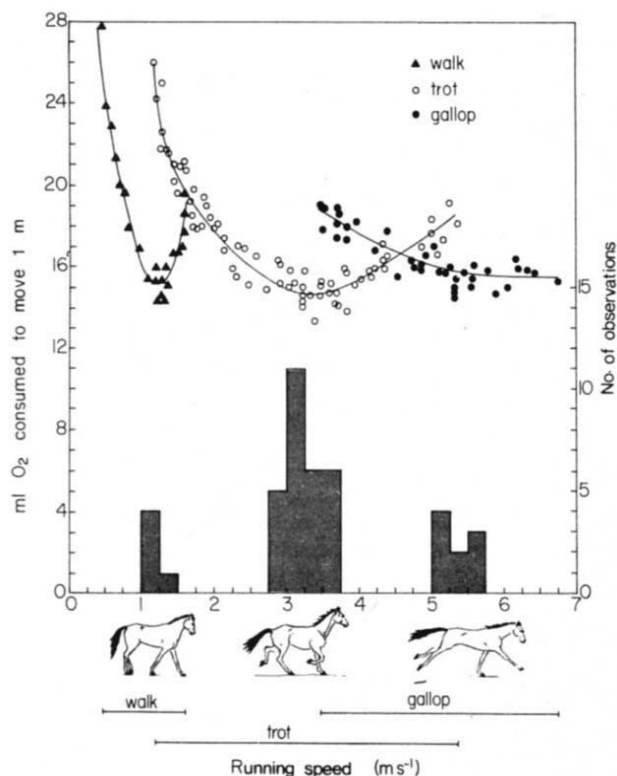


Fig. 2 The oxygen cost to move a unit distance (rate of oxygen consumption divided by speed, calculated from data in Fig. 1) declined to a minimum and then increased with increasing speed in a walk and trot. It also declines to a minimum in a gallop, but we were unable to obtain sufficiently high galloping speeds to observe any increase at high galloping speed if it occurred. The minimum oxygen cost to move a unit distance was almost the same in all three gaits. The histogram shows gaits where horse B was allowed to select her own speed while running on the ground. She chose three speeds which coincided with the energetically optimal speed for each gait. On a motorized treadmill, the animal must move at the speed of the tread, but note that when running on the ground, there were ranges of speeds which the animal never used for any sustained period.

can only be observed when great care is taken to avoid any source of variability in the data. Therefore, it seems useful to continue to use a single linear relationship between rate of oxygen consumption and speed for comparative studies of energetics of locomotion.

We conclude that horses, like humans, change gait and select speed within a gait in a manner that minimizes energy consumption. Observations of speeds used by migrating African animals³ suggest that this finding may also apply generally to terrestrial animals.

This work was supported by NIH grant 2R01 AM18140-04, -05, -06 and NIH postdoctoral fellowship 5 F32 AM05658-01, -02.

Received 19 February; accepted 29 April 1981.

1. Margaria, R. *Atti. Accad. naz. Lincei R.* 6, 7 (1938).
2. Knuttgen, H. G. *Acta physiol. scand.* 52, 366-371 (1961).
3. Pennycook, C. J. J. *exp. Biol.* 63, 775-799 (1975).
4. Seeherman, H. J., Taylor, C. R., Maloij, G. M. O. & Armstrong, R. B. *Resp. Physiol.* (in the press).

Gene expression in visna virus infection in sheep

Michel Brahic, Linda Stowring, Peter Ventura & Ashley T. Haase

Infectious Disease Section, Veterans Administration Medical Center and Departments of Medicine and Microbiology, University of California, San Francisco, San Francisco, California 94121, USA

Visna is a slow degenerative disease of the central nervous system (CNS) of sheep caused by a nontumorigenic retrovirus¹. During the course of this disease, visna virus establishes a persistent infection of the CNS, lung and haematopoietic system, despite a specific humoral and cellular immune response. We have studied visna virus life cycle at the single-cell level in choroid plexus of experimentally infected animals, using a very sensitive and quantitative *in situ* hybridization assay². We report here that although proviral DNA is synthesized in significant amounts, its expression is blocked at the transcriptional level. This restriction of proviral DNA transcription offers an explanation for the slowness of the disease and the persistence of the infection.

Two aspects of visna infection are relevant to the question of slowness and persistence. First, the virus is predominantly cell-associated, especially in the blood; second, the amount of virus present at any time during the course of the disease is minimal³. This limitation in virus growth is not a consequence of the immune response because it is observed before virus-specific antibodies can be detected, and because immunosuppression of infected animals does not lead to a higher titre of virus⁴. Restricted virus replication therefore seems to be an intrinsic property of the virus-cell interaction in the tissues of the animal (*in vivo*). This *in vivo* situation is in sharp contrast to virus replication in ovine cells grown *in vitro*, where the virus grows to high titre and causes a characteristic cytopathic effect that culminates in cell death 48-72 h after infection.

The first clue to the mechanism of restriction of replication *in vivo* was revealed by studies which showed that proviral DNA can be detected by *in situ* hybridization in a significant fraction

of the cells in the choroid plexus, but that only an occasional cell contains the major structural viral polypeptide p30 as detected by immunofluorescence⁵. This restriction of proviral DNA expression provided an explanation for the persistence of the infection: the immune defence mechanisms are unable to eradicate the infection because the virus persists intracellularly in cells that do not express viral antigens. Therefore, an understanding of the mechanism of restriction of viral gene expression is central to the study of visna pathogenesis.

We chose to study viral replication in the choroid plexus because histopathological lesions of visna always predominate in this tissue as well as in the periventricular areas of the brain³, and because persistent infection of the choroid plexus has been clearly documented by isolation studies³. Although it is a relatively simple structure, the choroid plexus consists of various cell types (cuboidal epithelium cells, fibroblasts, endothelial cells and smooth muscle cells) interspersed with large numbers of inflammatory cells characteristic of visna lesions. Because of this complex setting and the possibility of local variations in the extent of infection, virus replication must be studied at the single-cell level. Viral protein synthesis was assessed by immunofluorescence using an indirect assay and the labelled avidin-biotin method of detection⁶. Virus nucleic acids were quantitated in single cells in histological sections by *in situ* hybridization² modified for the detection of proviral DNA as described in Fig. 1 legend. This assay will detect 1 to 2 copies of viral RNA and 10 copies of proviral DNA per cell after three weeks of autoradiographic exposure (10 grains per cell above background).

Three American lambs were inoculated intracranially with 10⁹ plaque-forming units (PFU) of visna virus strain 1514. The inoculum (1 ml) was deposited in the vicinity of the lateral ventricle. A large inoculum was chosen to infect a maximum number of cells at the site of inoculation. The animals were killed between 7 and 15 days after inoculation, and the choroid plexi removed. Inflammatory lesions typical of visna were observed in all cases. A portion of choroid plexus was frozen, subsequently sectioned in a cryomicrotome and assayed for viral nucleic acids and proteins. Examples of the results obtained by *in situ* hybridization are shown in Fig. 2. Proviral DNA and viral RNA were readily detected in these sections. In all cases where

the cells could be unambiguously identified, viral nucleic acids were present only in cuboidal epithelium cells; they were not found in vessels, connective tissue cells or inflammatory mononuclear cells.

The fraction of choroid plexus cells carrying proviral DNA and viral RNA was determined (Table 1); values range between

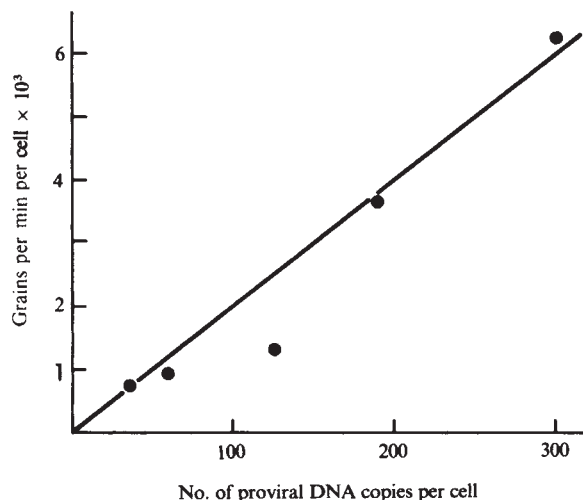


Fig. 1 Relationship between the amount of proviral DNA per cell and the number of autoradiographic grains developed after *in situ* hybridization. Sheep choroid plexus cells grown *in vitro* were infected with 3 PFU of visna virus per cell. At various times after infection (10, 14, 23, 48 and 72 h), the cells were collected and subjected in parallel to *in situ* hybridization and determination of proviral DNA content by liquid hybridization⁹. For *in situ* hybridization microscope slides were treated with bovine serum albumin, Ficoll and polyvinyl pyrrolidone as described elsewhere². The cells were deposited on treated slides using a cytocentrifuge, air dried for a few minutes and fixed for 4 min at -20°C in a mixture of methanol and acetone (1:2, v/v). The slides were treated with HCl, heat and proteinase K as described elsewhere². RNA was eliminated by treating the slides for 30 min at 37°C with $100\text{ }\mu\text{g ml}^{-1}$ of RNase A and 10 units ml^{-1} of RNase T₁ in $2\times\text{SSC}$ (0.3 M NaCl, 0.03 M Na citrate, pH 7.5), followed by two washes in $2\times\text{SSC}$ and dehydration in graded ethanol. To ensure maximal retention of viral DNA, the slides were refixed in 5% formaldehyde at room temperature for 2 h, then washed twice in $2\times\text{SSC}$ and dehydrated in graded ethanol. DNA denaturation was achieved at 65°C for 15 min in 95% deionized formamide, $0.1\times\text{SSC}$ followed by quenching in ice-cold $0.1\times\text{SSC}$. The slides were washed twice more in $0.1\times\text{SSC}$ and dehydrated in graded ethanol. Hybridization *in situ* was as described elsewhere² except that the concentration of cDNA was $0.8\text{ ng }\mu\text{l}^{-1}$. Washing of the slides, dipping in NTB-3 Kodak emulsion and exposure was as described elsewhere². The slides were exposed for 3 weeks (10-h infected cells), 11 days (14 and 24-h infected cells) and 48 h (48 and 72-h infected cells). An average of 1,500 grains were counted for each slide; the extent of *in situ* hybridization was expressed as number of grains per min of exposure time per nucleus. Uninfected cells were processed in parallel for the determination of background. The background was 1.7×10^{-4} grains per min per nucleus.

1 and 3%. For a given animal the same fraction of cells contained proviral DNA and viral RNA. Table 1 also shows averaged values of proviral DNA and viral RNA content per cell. The distribution of values for individual cells is shown in Fig. 3. On average, infected cells contained 60–70 copies of proviral DNA per cell. Although the values are spread over a relatively broad range, cells containing levels of proviral DNA comparable to those observed at the end of the permissive cycle (200–300 copies per cell (B. Traynor, personal communication)) were extremely rare (Fig. 3). The average content of viral RNA in the tissues varied between 100 and 180 copies of 35S RNA genome equivalents, and the distribution of RNA content among cells was relatively narrow, with almost no cells containing more than 500 copies. These values are almost two orders of magnitude below the level of RNA synthesis observed at the end

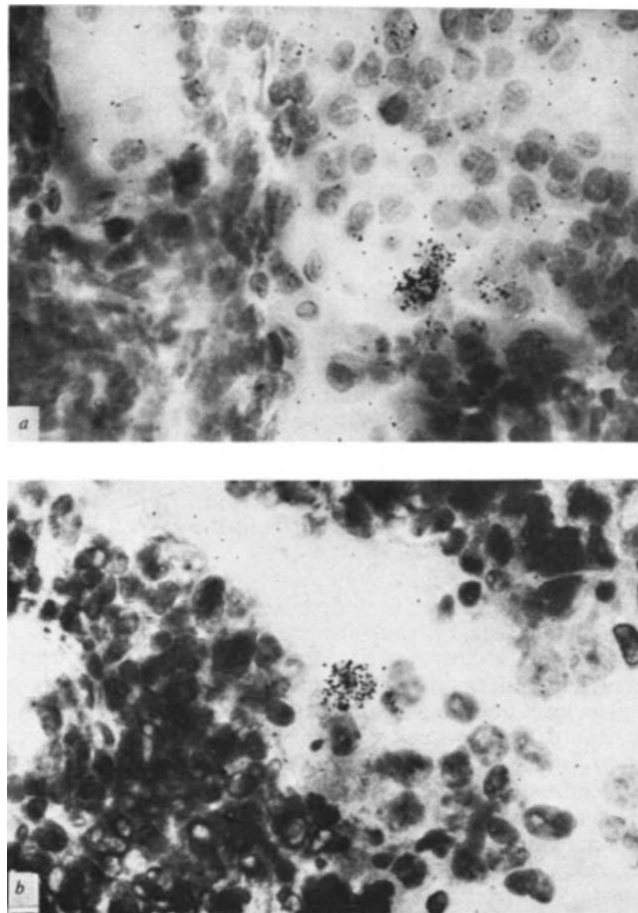


Fig. 2 Detection of proviral DNA (a) and viral RNA (b) in sections of choroid plexus from experimentally infected sheep. The figure shows one representative field in each case. Exposure times were 3 weeks (a) and 48 h (b).

Table 1 Numbers of choroid plexus cells carrying proviral DNA and viral RNA

Sheep no.	Days after inoculation	Viral DNA No. of positive cells	Viral RNA No. of positive cells	Average no. of copies per cell*		% Of cells containing viral proteins
		Total no. of cells (%)	Total no. of cells (%)	Viral DNA	Viral RNA	
927	7	142/1.5 $\times 10^4$ (0.9)	325/2.5 $\times 10^4$ (1.3)	61	177	0
1	10	156/1.5 $\times 10^4$ (1.0)	394/2.5 $\times 10^4$ (1.6)	74	129	0.0025
924	15	429/1.5 $\times 10^4$ (2.8)	684/2.5 $\times 10^4$ (2.7)	61	105	0

The number of cells positive for viral DNA was measured after 6 weeks of autoradiographic exposure, and for viral RNA was measured after 3 weeks of exposure. Total number of cells was estimated from the number of cells within the area of a reticle and the ratio of the area of the reticle to that of the section. Percentage of cells containing viral proteins was determined by immunofluorescence⁶ using anti-p30 serum.

* One copy corresponds to a MW of 6×10^6 double-stranded proviral DNA or 3×10^6 single-stranded viral RNA.

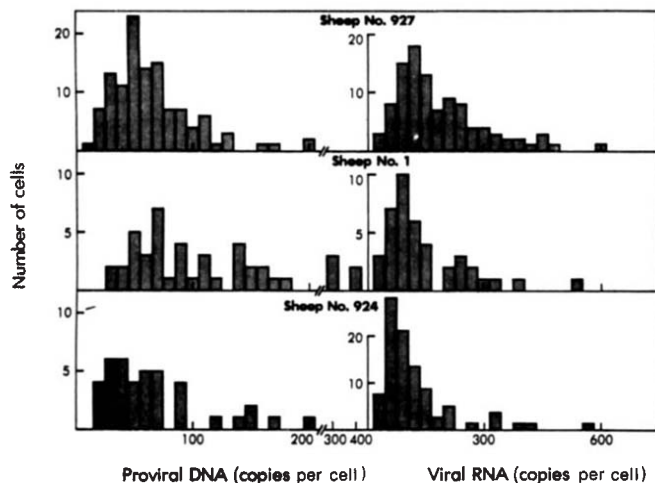


Fig. 3 Distribution of proviral DNA and viral RNA content among choroid plexus cells *in vivo*. The choroid plexus of three experimentally infected animals were examined by *in situ* hybridization. The autoradiographic exposure times were 3 weeks and 48 h for proviral DNA and viral RNA detection, respectively. In each case a representative section ($\sim 12 \text{ mm}^2$) was systematically scanned under the microscope at a magnification of $\times 1,000$. The number of grains per cell was counted for each positive cell. The number of genome copies per cell was determined from calibration curves (see Fig. 1 and ref. 2). The histograms were constructed by grouping copy numbers to the nearest multiple of 10.

of a permissive cycle ($\approx 10,000$ copies per cell⁷). Hybridization to proviral DNA was totally dependent on denaturation and was sensitive to pretreatment with DNase, therefore ruling out the possibility that the hybrids observed in the DNA assay were due to residual viral RNA. Immunofluorescence was carried out on subjacent sections of choroid plexus from the same animals; no convincingly positive cells were detected in two animals. A positive cell was occasionally found in the choroid plexus of sheep no. 1 (Table 1).

This quantitative analysis of viral replication in the choroid plexus demonstrates the following: (1) visna virus infects cuboidal epithelial cells. (2) Despite a relatively large inoculum deposited in the vicinity of the choroid plexus, the proportion of cuboidal choroid plexus cells infected is relatively low ($\approx 1\%$). (3) The level of viral RNA expressed in these cells is minimal. Infected cells constitute a fairly homogeneous population of cells containing 100–200 viral genome equivalents. In this respect these cells are similar to *in vitro* infected permissive cells arrested at a very early stage of the viral life cycle. Although they contain relatively high levels of proviral DNA (60–70 copies per cell), their RNA content is about two orders of magnitude lower than during a permissive viral replication cycle. This shows that the restriction of proviral DNA expression is regulated, at least in part, at the transcriptional level. (4) Choroid plexus cells do not contain viral structural proteins as determined by immunofluorescence.

These results represent a new step in the understanding of visna pathogenesis. Two characteristic aspects of this disease are the low level of virus expression and the persistence of the infection for long periods of time in spite of an immune response. Here we have documented the restriction of virus expression at the single-cell level. Minimal virus production is due to the limitation of the infection to a small fraction of the cells and the existence in these cells of a block at the level of proviral DNA transcription. We propose that restricted proviral DNA expression plays a part in two aspects of visna pathogenesis. First, it provides a damping mechanism which prevents the rapid spread of the infection and its expression as an acute disease and also limits the damage caused by viral replication to individual cells. (Indeed cell necrosis is conspicuously absent in visna, particularly during the early incubation period⁸.) Second, it allows the virus to persist: cells which are not killed as a result of the infection and which do not express viral antigens at their

surface provide the virus with a site where it can persist without being detected and eradicated by the immune defence mechanisms.

This work was supported by the Veterans Administration and by N I Health grant NS11782 and American Cancer Society grant MV50F. A.T.H. is supported by a Veterans Administration Medical Investigatorship.

Received 22 December 1980; accepted 27 March 1981.

1. Brahic, M. & Haase, A. T. in *Comparative Diagnosis of Viral Diseases* Vol. 3 (ed. Kurstak, E.) (Academic, New York, in the press).
2. Brahic, M. & Haase, A. T. *Proc. natn. Acad. Sci. U.S.A.* **75**, 6125–6129 (1978).
3. Petrusson, G., Nathanson, N., Georgsson, G., Panitch, H. & Palsson, P. A. *Lab. Invest.* **35**, 402–412 (1976).
4. Panitch, H., Petrusson, G., Georgsson, G., Palsson, P. A. & Nathanson, N. *Lab. Invest.* **35**, 452–560 (1976).
5. Haase, A. T., Stowring, L., Narayan, O., Griffin, D. & Price, D. *Science* **195**, 175–177 (1977).
6. Guesdon, J., Ternynck, T. & Avrameas, S. *J. Histochem. Cytochem.* **27**, 1131–1139 (1979).
7. Brahic, M., Filippi, P., Vigne, R. & Haase, A. T. *J. Virol.* **24**, 74–81 (1977).
8. Georgsson, G., Palsson, P. A., Panitch, H., Nathanson, N. & Petrusson, G. *Acta neuropath.* **37**, 127–135 (1977).
9. Haase, A. T. & Varmus, H. E. *Nature new biol.* **245**, 237–239 (1973).

Apparent reversion of stable *in vitro* genetic markers detected in tumour cells from spontaneous metastases

J. Dennis, T. Donaghue, M. Florian & R. S. Kerbel

Cancer Research Laboratories, Department of Pathology, Queen's University, Kingston, Ontario, Canada K7L 3N6

The evolution towards more aggressive and autonomous behaviour of many cancerous tumours, often referred to as tumour progression¹, is thought to stem from the development of heterogeneity within the tumour cell population, combined with the continuous selection of progressively more malignant cellular phenotypes^{2,3}. During the course of the disease, the tumour cells show multiple phenotypic changes in a stepwise, but apparently random fashion, becoming more anaplastic, increasingly independent of growth controls and more metastatic⁴. Several laboratories, including our own, have analysed aspects of tumour heterogeneity and cancer metastasis by selecting and studying the properties of lectin-resistant (Lec^R) membrane mutant tumour sublines^{5–8}; in a few cases, such variants have been claimed to be less tumorigenic^{5,6} or metastatic^{7,8} than the parental cells from which they were derived. We have attempted to study the factors involved in the re-establishment of tumour heterogeneity by monitoring the stability *in vivo* of Lec^R phenotypes of metastatic tumour cells after injection of cloned Lec^R tumour cells. We now report that spontaneous metastases arising after a subcutaneous (s.c.) injection of cells from variant tumour lines selected from a highly metastatic DBA/2 mouse tumour known as MDAY-D2⁹, and which were stably resistant in tissue culture to wheat germ agglutinin (WGA), no longer carry the WGA-resistant (WGA^R) phenotype. The results demonstrate that WGA^R tumour cells do not metastasize, but rather, 'revertants' for the WGA^R phenotype, which presumably were generated *in vivo* after injection, were the cells actually capable of metastatic growth.

The cloned WGA^R variant subline MDW4 was selected from the parental MDAY-D2 tumour line by its ability to grow in normally toxic concentrations (for example, 20–50 $\mu\text{g ml}^{-1}$) of WGA⁵. We have previously shown that MDW4 carries a second phenotypic marker—a deficiency of fucose salvage (fucose salvage⁻) caused by a block in the conversion of fucose-1-P to GDP-fucose⁸. These phenotypes seemed to be remarkably stable in MDW4 cells maintained in tissue culture for more than 2.5 yr (Fig. 1). A second variant cell line, MDOW-2b, was sequentially selected for its ability to grow in high concentrations of the toxic membrane-active drug ouabain⁹ (3 mM)

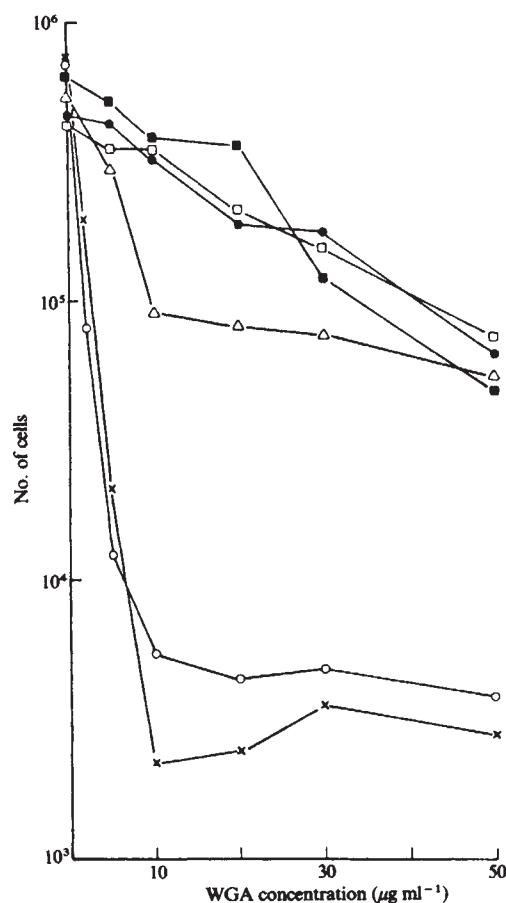


Fig. 1 Stability of the WGA^R phenotype of the MDW4 tumour clone maintained in tissue culture. Tumour cells were grown in tissue culture in α -MEM (Flow) plus 10% fetal calf serum with subculturing every 3–5 days. The variant was periodically tested for the ability to grow in the presence of WGA; Δ , 10 July 1978; $\bullet$, 6 December 1978; $\blacksquare$, 10 June 1980; $\square$, 1 October 1980. MDW4 cells were seeded at 3×10^3 cells ml⁻¹ in the presence of 0, 2, 5, 10, 20, 30 and 50 μ g ml⁻¹ WGA and were grown *in vitro* until those seeded in the absence of WGA had reached confluence (7–9 days), after which cell numbers were determined in each culture. The revertant, WGA-sensitive (WGA^S) phenotype of the tumour lines obtained from metastases and maintained in tissue culture also remained unchanged for the revertant phenotypes after more than 6 months. MDW4-22a cells were tested on 22 August 1980 ($\circ$) and again on 10 January 1981 ($\times$). The WGA^R profile of the parental MDAY-D2 cells is shown in Fig. 2, and has also been described previously⁵.

and WGA (20 μ g ml⁻¹). MDOW-2b did not possess the fucose salvage⁻ phenotype.

As reported previously⁸, WGA^R variants of MDAY-D2 grew and metastasized in an altered way in the syngeneic DBA/2 host. Thus, a s.c. injection of 10^5 MDAY-D2 tumour cells produced visible foci of tumour growth in the liver within 2 weeks of injection. In contrast, no liver metastases were evident 2 weeks after a s.c. injection of 10^5 cells of either MDW4 or MDOW-2b. The number of metastatic nodules, at death, in the livers of mice which had received 10^5 MDW4 cells s.c. ranged from 0 to 90, whereas mice which died after injection with the same dose of MDAY-D2 cells had almost complete liver replacement with tumour tissue (that is, the number of tumour foci were excessive and not discrete, and therefore not easily counted)⁸. An intravenous (i.v.) injection of 10^5 MDW4 cells results in no tumour takes, whereas i.v. injection of as few as 10 MDAY-D2 cells results in widespread metastases, similar to that observed after s.c. injections. Mice succumbed to the WGA^R variants in ~60 days, while the parental tumour MDAY-D2 killed the mice in 35 days (Table 1). We concluded from these results that the WGA^R variants, although still malignant, were less so than the parental MDAY-D2 tumour⁸.

Due to the potential ability of tumour cells to change phenotypes and become progressively more malignant in animals¹⁰, it

was not immediately clear whether the few metastases produced by a s.c. injection of the WGA^R variants was indicative of an inherently less metastatic tumour line, or of an essentially non-metastatic line which, as a result of the generation and selection of more malignant variants *in vivo*, showed limited metastasis of these new variants. If ouabain-resistant (Oua^R), WGA^R and/or fucose salvage⁻ phenotypes were related to a reduction in the metastatic potential of the tumour cells, then from the first hypothesis we could predict that metastases would retain the same phenotypes as the injected cells. Conversely, the latter hypothesis would predict that metastases would show a high incidence of revertant phenotypes. To examine this question, solid tumours at the site of injection and individual spontaneous metastases from DBA/2 mice which had been injected with either MDW4 or MDOW-2b cells were removed and re-established as tissue culture lines. The tumour cells were removed from single, discrete metastatic foci obtained from liver, kidney, brain, spleen or intestine.

Tumour cells at the site of injection and metastases resulting from a s.c. injection of MDOW-2b tumour cells retained their Oua^R phenotype but were revertant for WGA^R (Fig. 2). Retention of the Oua^R phenotype indicated that the tumour nodules originated from the injection of MDOW-2b cells and were not new tumours, induced *in situ*¹¹. Similarly, all metastases resulting from s.c. MDW4 cells were revertant for the WGA^R and fucose salvage⁻ phenotypes, two characteristics of tumour cells which have previously been suggested to influence metastatic capacity^{6–8,12} (Table 1). The reverted phenotype of the tumour lines derived from metastases were stable as they remained unchanged after more than 6 months in tissue culture (Fig. 1). Furthermore, injection of the cultured tumour cells established from metastatic foci resulted in rapid tumour growth (27–35 days for 50% mortality) and the same metastatic pattern of liver replacement observed with the parental tumour MDAY-D2 (Table 1).

In contrast to the metastases, not all the 'primary' s.c. growing solid tumours were complete revertants for both WGA^R and fucose salvage⁻ (Table 1). An inoculum of 10^2 MDW4 cells produced a solid tumour which was a 'complete revertant' (MDW4-11) while inocula of 10^4 and 10^5 MDW4 cells produced tumour cell populations at the site of injection which seemed to be only partially reverted for either the WGA^R or fucose salvage⁻ phenotypes. The finding that injection of a low dose of MDW4 cells resulted in a solid tumour which was a complete revertant (MDW4-11), while higher doses of MDW4 produced tumours at the site of inoculation which were only partially reverted for the WGA^R and fucose salvage⁻ phenotypes (a heterogeneous population of cells), suggests, but does not prove, that the revertants did not exist in the original inoculum of MDW4 cells, but arose later in the animals and were selected as a consequence of their superior ability to survive *in vivo*. Hence, it seemed that the MDW4 metastases were revertant for both WGA^R and fucose salvage⁻, even in the cases where the solid tumour displayed only a partially reverted phenotype; this suggests that the cells capable of metastasizing consisted of a small subpopulation of tumour cells originating at the site of inoculation and displaying revertant phenotypes.

The results of the following experiment support this interpretation (Table 2). Intravenous injection of MDW4 cells, as already stated, even when as many as 10^5 cells are given, does not result in any tumour takes (Table 2). The same is true when nude mice are used as recipients (data not shown). However, if one mixes as few as 10 cells from a metastasis (designated MDW4-22a)—which arose in a DBA/2 mouse after s.c. injection of MDW4 cells—with 10^5 MDW4 cells, and injects the mixture i.v., 75% ($n = 4$) of the animals rapidly develop metastases and die within 4 weeks. In view of these results and the fact that the s.c. injection of as few as 100 MDW4 cells leads to WGA-sensitive metastases (Table 1), it is difficult to argue that pre-existing cells revertant for the WGA^R phenotype in the MDW4 cell line were entirely responsible for the reported findings.

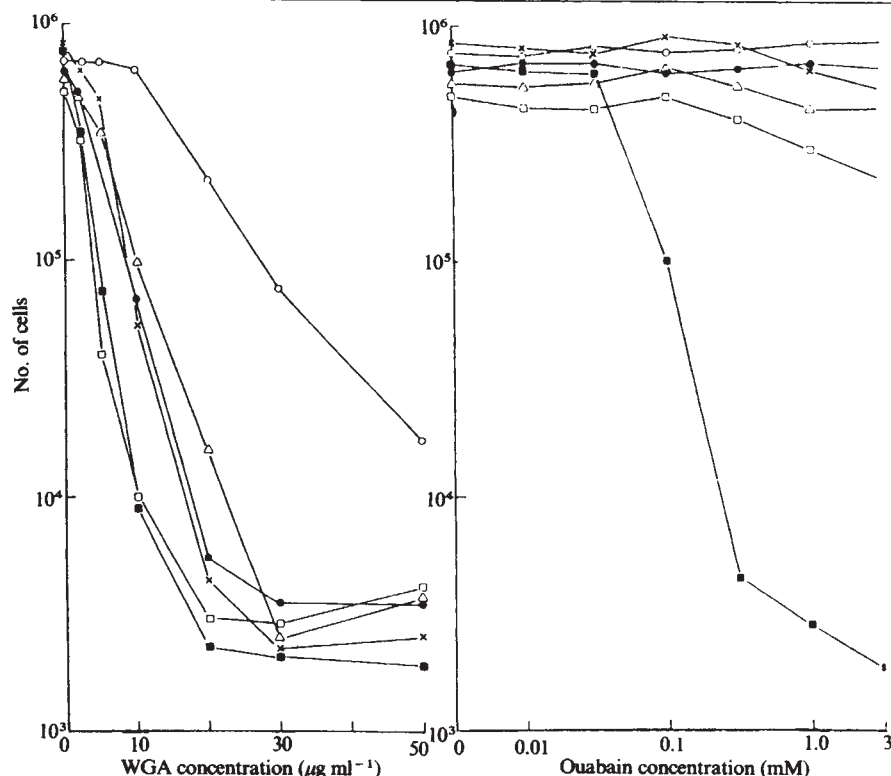


Fig. 2 WGA and ouabain sensitivity of MDAY-D2, MDOW-2b and of cell lines established from spontaneous metastases removed from the liver of an MDOW-2b-bearing mouse. DBA/2 mice were given a s.c. injection of 10⁵ MDOW-2b cells followed 60 days later by removal of tumour tissue from the site of injection (MDOW-11) and three clearly discrete metastatic nodules found in the liver (MDOW-21a, MDOW-21b, MDOW-21c). Tumour tissue removed from the mouse was minced, trypsin treated and the resulting cell suspension placed in tissue culture for at least 3 weeks to allow tumour cells to outgrow host-derived cells. The established tissue culture lines were then tested for the ability to grow in the presence of either WGA (left panel) or ouabain (right panel) as described in Fig. 1 legend; ■, MDAY-D2; ○, MDOW-2b; ×, MDOW-11; △, MDOW-21a; □, MDOW-21b; ●, MDOW-21c.

Several mechanisms can be envisaged for the generation of the revertant and more malignant phenotypes—classical genetic mechanisms of back mutation, acquisition or loss of genetic material by the tumour cells, stable regulatory changes in gene expression or somatic cell hybridization *in vivo* between a tumour cell and a host cell. The last possibility has been demonstrated previously¹³ but seemed unlikely in the present situation because *Oua*^R is a co-dominant marker⁹, such that a somatic

hybrid between *Oua*^R and a wild-type cell results in a partial loss of *Oua*^R in the resultant hybrid⁹, an event not observed for the metastases of MDOW-2b. We are now carrying out experiments to examine this possibility in more detail, and the other potential mechanisms involved in the generation of the revertant phenotypes *in vivo*.

The results demonstrate a need to examine the progeny, especially metastases, arising from an injection of selected

Table 1 WGA^R and fucose salvage⁻ phenotypes of subcutaneous and metastatic tumour nodules

Cell line	Source	WGA resistance (mg ml ⁻¹)	³ H-fucose/ ³ H-leucine* (c.p.m./c.p.m. × 10 ⁻¹)	50% mortality† (days)	Liver metastases at death‡
MDW4-11	1°, 10 ² s.c.	2	1.70 ± 0.30 (195)	33	C
MDW4-21a	2°, liver	2	Not tested	30	C
MDW4-21b	2°, liver	2	1.91 ± 0.56 (220)	35	C
MDW4-21c	2°, liver	2	1.70 ± 0.28 (195)	32	C
MDW4-12	1°, 10 ⁴ s.c.	2	0.38 ± 0.07 (44)	29	C
MDW4-22a	2°, liver	2	0.73 ± 0.18 (84)	30	C
MDW4-22b	2°, liver	2	0.70 ± 0.35 (81)	27	C
MDW4-22c	2°, kidney	2	0.84 ± 0.16 (97)	31	C
MDW4-22d	2°, brain	2	0.71 ± 0.07 (82)	30	C
MDW4-14	1°, 10 ⁴ s.c.	2	0.08 ± 0.03 (9)	32	C
MDW4-24a	2°, liver	2	1.16 ± 0.11 (133)	34	C
MDW4-24b	2°, liver	2	2.05 ± 0.89 (236)	34	C
MDW4-24c	2°, liver	2	1.84 ± 0.37 (211)	35	C
MDW4-13	1°, 10 ⁵ s.c.	43	0.26 ± 0.1 (30)	30	C
MDW4-23a	2°, liver	2	1.37 ± 0.25 (157)	31	C
MDW4-23b	2°, liver	2	1.27 ± 0.28 (146)	29	C
MDW4-23c	2°, spleen	2	1.18 ± 0.09 (136)	33	C
MDW4-23d	2°, gut	2	1.33 ± 0.17 (153)	30	C
MDAY-D2	Tissue culture line	5	0.87 ± 0.21 (100)	35	C
MDW4	Tissue culture line	50	0.03 ± 0.01 (3)	63	N(3-90)

DBA/2 mice were given a s.c. injection of the WGA^R, fucose salvage⁻ tumour cells MDW4. Seven to ten weeks after the injection of 10², 10⁴ or 10⁵ MDW4 cells, tumour tissue at the site of injection (1°) and three or four secondary metastatic nodules (2°) were removed, minced with scissors, trypsin treated to make a cell suspension and established as tissue culture lines. The cell lines derived from metastases (2°), which are listed below each primary (1°), were removed from liver, lung, spleen, brain, kidney or intestine of the same mouse. The concentration of WGA necessary to cause a 90% inhibition of tumour cell growth in tissue culture was determined as described in Fig. 1 legend.

* Tumour cells, 10⁵ in 200 μl of culture medium, were pulse-labelled with 1 μCi of L-[6-³H]fucose (27 Ci mmol⁻¹) and with 1 μCi of L-[4, 5-³H]leucine (60 Ci mmol⁻¹) in parallel for 16 h at 37 °C. Acid-precipitable material was washed three times in 10% trichloroacetic acid followed by solubilization and determination of radioactivity. The ratio of fucose to leucine was determined for each cell line and the numbers in parentheses are ³H-fucose to ³H-leucine ratios expressed as a percentage of the ³H-fucose to ³H-leucine ratio for MDAY-D2. The results are the mean of three determinations performed in duplicate ± s.d.

† Twelve DBA/2 mice were given s.c. injections of 10⁵ MDW4 or MDAY-D2 cells and the number of days required for 50% of the animals to succumb to the tumours was recorded. For each of the 1° and 2° tumour cell lines, four DBA/2 mice were used for the 50% mortality determinations. The tumour lines were maintained in tissue culture for at least a month before assessing their malignancy.

‡ Tumour infiltration of the liver was examined at death and described as confluent with tumour (C) or infiltrated with a limited number of nodules (N); the range is indicated in parentheses.

Table 2 Tumour takes in DBA/2 mice injected intravenously with MDW4 and cells from MDW4 metastases

Tumour cells injected	Tumour takes per no. of mice injected
MDW4 (10^5)	0/8
MDW4 (10^5) + MDW4-22a (10)	3/4
MDW4 (10^5) + MDW4-22a (10^2)	4/4
MDW4-22a (10^2)	4/4

DBA/2 mice were given a single i.v. injection of MDW4, MDW4-22a or a mixture of the two cell lines, and tumour growth was determined at death by autopsy. Animals with tumour died 25–35 days after tumour injection. Those that survived were found to be tumour free 4–5 months after inoculation. MDW4-22a refers to a cell line established from a liver metastasis which had been removed from an animal previously injected s.c. with MDW4 cells. The numbers in parentheses refer to the number of cells injected i.v.

tumour variants to determine whether or not they are phenotypically the same as the cells injected, especially when the selecting agent seems to influence the malignant potential of the variant (WGA). Tumour variants selected *in vitro* which display a decreased malignant potential are apparently subject to the forces which generate tumour heterogeneity *in vivo*. Consequently, revertant and more malignant phenotypes may

arise relatively quickly when the cells are injected into animals. Our work extends the results of Tao and Burger⁶, who have shown that WGA^R variant tumour cells are often poorly metastatic. We have demonstrated that the WGA^R variants may be non-metastatic and that the WGA^R tumour cells must undergo phenotypic alterations, including a reversion of WGA^R, before metastasis can occur.

We thank Drs Louis Siminovitch and Victor Ling for helpful discussions and Mrs B. Fluhrer and Mrs Janie Barr for secretarial assistance. This work was supported by the NCI of Canada, of which R.S.K. is a Research Associate.

Received 11 February; accepted 6 May 1981.

1. Prehn, R. T. *Adv. Cancer Res.* **23**, 203–236 (1976).
2. Nowell, P. C. *Science* **14**, 23–28 (1976).
3. Klein, G. & Klein, E. *Transplant Proc.* **9**, 1095–1104 (1977).
4. Poste, G. & Fidler, I. J. *Nature* **283**, 139–146 (1980).
5. Kerbel, R. S. *Am. J. Path.* **97**, 609–622 (1979).
6. Tao, W.-W. & Burger, M. M. *Nature* **270**, 437–438 (1977).
7. Reading, C. L., Belloni, P. N. & Nicolson, G. J. *J. natn. Cancer Inst.* **64**, 1241–1249 (1980).
8. Dennis, J. W. & Kerbel, R. S. *Cancer Res.* **41**, 98–104 (1981).
9. Baker, R. M. *et al. Cell* **1**, 9–21 (1974).
10. Schirmacher, V. *Immunobiology* **157**, 89–98 (1980).
11. Kerbel, R. S., Florian, M., Man, M. S., Dennis, J. & McKenzie, I. F. C. *J. natn. Cancer Inst.* **64**, 1221–1231 (1980).
12. Vischer, P. & Reutter, W. *Eur. J. Biochem.* **84**, 363–368 (1978).
13. Wiener, F., Fenyo, E. M. & Klein, G. *Nature new Biol.* **238**, 155–159 (1972).

A single dominant gene determines susceptibility to a leukaemogenic recombinant retrovirus

Robert S. Schwartz & Raman H. Khirya

Department of Medicine, Tufts University School of Medicine, Boston, Massachusetts 02111, USA

The production of recombinant retroviruses is an important episode in the natural history of thymic leukaemia in AKR and HRS/J (*hr/hr*) mice^{1,2}. These viruses apparently originate from ecotropic and xenotropic precursors in the late preleukaemic stage of the disease. Analyses of their structural proteins^{3,4} and genomic oligonucleotides^{3,5} indicate that they arise by recombination of *env* genes of the precursor viruses. This event leads to a viral envelope glycoprotein (gp70) with some peptides that have features of the gp70 glycoproteins of ecotropic and xenotropic viruses, and others that are unique for each recombinant virus^{3,4,6}. The former property explains the broad host range of recombinant viruses, and hence their designation as dual tropic¹ or polytropic² viruses. It has been postulated that the unique aspect of each recombinant's gp70 determines the phenotypes of leukaemic cells⁷. Polytropic viruses may be highly thymotropic. Their systemic administration results in an infection that confines itself virtually to the thymus^{3,8}. Moreover, these viruses are leukaemogenic whereas their precursors are not, or only weakly so^{3,8,9}. The leukaemogenicity of polytropic viruses is, however, restricted to certain inbred strains of mice^{3,8}. The HRS/J isolate PTV-1 is leukaemogenic in HRS/J and CBA/J mice, but not in SWR/J or NIH/Swiss mice³. The experiments described here demonstrate that a single dominant gene permits infection of thymocytes by a leukaemogenic polytropic virus. CBA/J mice, which develop thymic leukaemia after infection by this virus, possess this gene, whereas leukaemia-resistant NFS mice lack it.

Figure 1 shows the results when newborn mice were inoculated with 10^3 – 10^4 focus-forming units (FFU) of a cloned polytropic virus, PTV-1. This thymotropic virus is leukaemogenic in CBA/J, but not in NIH/Swiss mice or their inbred counterpart, NFS (ref. 3 and unpublished observations). Its gp70 and genomic RNA are typical of a recombinant virus³. The mice were killed 4–6 weeks after inoculation and the levels of polytropic virus in their thymuses were measured by an infectious centre assay¹⁰. Polytropic virus is undetectable in uninoculated animals of the strains used in these experiments³.

CBA/J mice were readily infected by PTV-1, whereas NFS mice had either low or undetectable levels of virus. F₁ mice behaved like CBA/J mice. Susceptibility to PTV-1 is therefore a dominant trait. The upper limit of resistance (100 FFU per 10^7 thymocytes) was defined by the mean + 2 s.d. of the results in NFS mice and the lower limit that defined susceptibility (440 FFU per 10^7 thymocytes) was taken as the mean – 2 s.d. of the results in CBA/J mice. The 66 backcrosses to the NFS parent segregated, by these standards, into susceptible and resistant mice in a ratio of 38:28. This ratio is not significantly different from 33:33 ($\chi^2 = 0.88$, $P > 0.1$). The corresponding ratio in 61 F₂ mice was 44:17, which is not significantly different from 45:15 (3:1) ($\chi^2 = 0.14$, $P > 0.1$). Exclusion of the eight F₂ mice with intermediate values (between 100 and 440 FFU per 10^7 thymocytes) does not change the significance of the result ($\chi^2 = 0.53$, $P > 0.1$). All backcrosses to the CBA/J parent were susceptible. These results are consistent with the conclusion that a single dominant gene confers susceptibility to infection by PTV-1. We shall refer to this gene as *PTV*¹.

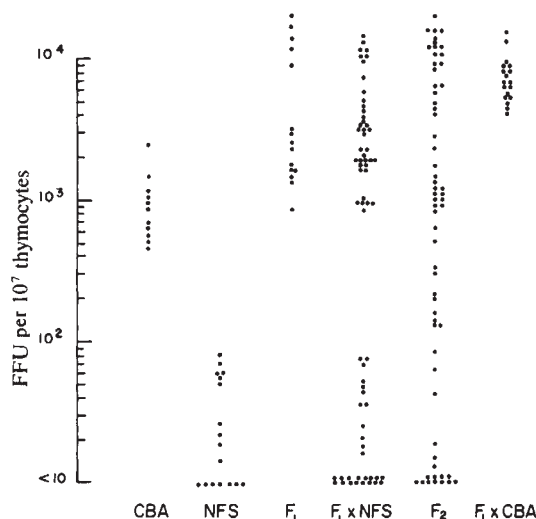


Fig. 1 Results of infection with PTV-1. Mice were injected intraperitoneally with 10^3 – 10^4 FFU of PTV-1 within 24 h of birth. The animals were killed 4–6 weeks later and cell suspensions prepared from their thymuses. These cells were tested by an infectious centre assay³ on mink lung cells (CCL64). X-C negative viruses capable of infecting both mink and mouse (NIH 3T3) cells were detected in all CBA/J mice injected with PTV-1. Neither mink-tropic nor mouse-tropic viruses can be detected in age-matched uninjected CBA/J or NFS mice.

Table 1 Results of infection with ETV-1 (cloned ecotropic retrovirus³)

	CBA	NFS
ETV-1	3.3 ± 0.7	3.5 ± 1.2
PTV-1	3.1 ± 0.4	0.9 ± 0.5

Mice were injected intraperitoneally with 10^3 – 10^4 FFU of ETV-1 within 24 h of birth. The animals were killed 4–6 weeks later and thymocytes were tested in an infectious centre X-C assay¹. Values for ETV-1 are expressed as $\log_{10}$ X-C plaques/ 10^7 thymocytes $\pm$ s.e. Data for PTV-1 are taken from Fig. 1.

The mechanism whereby PTV¹ confers susceptibility to PTV-1 is unknown. Its action is specific for the polytropic virus, because NFS mice are as susceptible as CBA/J mice to infection by a cloned ecotropic virus (Table 1). The effect of PTV¹ does not seem to be immunological because CBA/J and NIH/Swiss

mice produced comparable levels of antiviral antibodies after neonatal infection with either PTV-1 or an ecotropic virus (unpublished observation). We favour the view that this gene specifies a receptor on thymocyte surface membranes for the constant region⁶ of recombinant gp70 because that glycoprotein is a major determinant of viral host range¹¹. The specificity of PTV¹ for the recombinant virus (Fig. 2) supports this interpretation, as the only known difference between viruses of this type and ecotropic viruses is in the gp70 (refs 3,4,6). We cannot exclude an effect that would be analogous to that of Fv-1, which acts after penetration of the virus¹². This possibility seems unlikely, however, because PTV-1 readily infects NIH 3T3 cells *in vitro*³. Finally, the present results support the view^{3,8} that the highly selective thymotropism of certain polytropic viruses is an important factor in their leukaemogenicity.

This work was supported by NIH grants CA10018 and CA24530.

Received 23 February; accepted 6 May 1981.

- Hartley, J. W., Wolford, N. K., Old, L. J. & Rowe, W. P. *Proc. natn. Acad. Sci. U.S.A.* **74**, 789–792 (1977).
- Hiai, H., Morrissey, P., Khirya, R. & Schwartz, R. S. *Nature* **279**, 247–249 (1977).
- Green, N. *et al. J. exp. Med.* **152**, 249–264 (1980).
- Elder, J. H. *et al. Proc. natn. Acad. Sci. U.S.A.* **74**, 4676–4680 (1977).
- Rommelaere, J., Faller, D. V. & Hopkins, N. *Proc. natn. Acad. Sci. U.S.A.* **74**, 495–499 (1978).
- Elder, J. H., Jensen, F. C., Bryant, K. L. & Lerner, R. A. *Nature* **267**, 23–28 (1977).

- Zielinski, C. C., Waksal, S. D., Tempelis, L. D., Khirya, R. H. & Schwartz, R. S. *Nature* **288**, 489–490 (1980).
- Cloyd, M. W., Hartley, J. W. & Rowe, W. P. *J. exp. Med.* **151**, 542–552 (1980).
- Nowinski, R. C. & Hayes, E. F. *J. Virol.* **27**, 13–18 (1978).
- Kawashima, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **74**, 789–792 (1976).
- Stephenson, J. R. in *Molecular Biology of RNA Tumor Viruses*, 245–297 (Academic, New York, 1980).
- Pincus, T. in *Molecular Biology of RNA Tumor Viruses*, 77–130 (Academic, New York, 1980).

Sequence and specificity of two antibacterial proteins involved in insect immunity

H. Steiner*, D. Hultmark*, Å. Engström†, H. Bennich† & H. G. Boman*

*Department of Microbiology, University of Stockholm, S-106 91 Stockholm, Sweden

†The Biomedical Centre, University of Uppsala, S-751 23 Uppsala, Sweden

Immune responses have been described for many different insect species¹. However, it is generally acknowledged that insects lack lymphocytes and immunoglobulins and their immune systems must therefore differ from those of vertebrates. An effective humoral immune response has been found in pupae of the cecropia moth, *Hyalophora cecropia*¹. The expression of this multicomponent system requires *de novo* synthesis of RNA and proteins² and its broad antibacterial activity is due to at least three independent mechanisms³, the most well known of which is the insect lysozyme^{4–7}. However, this enzyme is bactericidal for only a limited number of Gram-positive bacteria. We recently purified and characterized P9A and P9B, which are two small, basic proteins with potent antibacterial activity against *Escherichia coli* and several other Gram-negative bacteria⁷. We believe that P9A and P9B play an important part in the humoral immune responses described previously⁸ and that the P9 proteins represent a new class of antibacterial agents for which we propose the name cecropins. We describe here the primary structures of cecropins A and B. We also show that cecropin A is specific for bacteria in contrast to melittin, the main lytic component in bee venom⁹ which lyses both bacteria and eukaryotic cells.

A two-step chromatographic purification of cecropins A and B has been recently described⁷. This purification was improved for sequence work by first using gel filtration on Sephadex G100.

The tentative sequences of both cecropins are given in Fig. 1. Residues 1–33 were obtained by automated Edman sequence analysis and the order of the C-terminal residues 34–37 was deduced from carboxypeptidase Y degradations. This part of the sequence is tentative until confirmed by an independent method. The C-terminal Lys and Ser residues could only be detected after acid hydrolysis of the carboxypeptidase digest; therefore we conclude that the C-terminals are blocked.

The cecropins have very similar overall structures. The N-terminal sequences (residues 1–10) are strongly basic and the central regions (residues 22–30) strongly hydrophobic. The amino acid compositions obtained from the sequences and from automated analyses are compared in Table 1. The agreement between the two sets of data support the structures shown in Fig. 1. Further evidence for the sequences was obtained by partial cleavage of cecropin A at residue 17 using acid hydrosol¹⁰ and isolation of the C-terminal fragment by high-voltage electrophoresis. The amino acid composition of this peptide was in agreement with the structure given in Fig. 1. Cecropin B was cleaved by cyanogen bromide at the single methionine in position 11 and the resulting fragments analysed. This analysis confirmed the order and identity of residues 1–33 but despite several different approaches, we have failed to obtain Edman sequencing data for residues 34–37.

Table 1 Amino acid compositions of cecropins A and B

Amino acid	Cecropin A		Cecropin B	
	Analysis	Sequence	Analysis	Sequence
Asp	2.01*	1	2.15*	0
Asn		1		2
Thr	1.07	1	0.33	0
Ser	0.17	0	0.53	1
Glu	3.89*	1	1.95*	2
Gln		3		0
Pro	0.95	1	1.04	1
Gly	4.05	4	4.07	4
Ala	4.88	5	4.84	5
Cys	0	0	0	0
Val	3.31	4	2.47	3
Met	0	0	0.80	1
Ile	3.65	5	3.48	5
Leu	1.09	1	1.95	2
Tyr	0	0	0.10	0
Phe	0.98	1	0.87	1
Lys	6.42	7	6.75	7
His	0.06	0	0.29	0
Arg	0.94	1	1.98	2
Trp	—	1	—	1
Total residues		37		37

Amino acid composition of the acid hydrolysate (6 M HCl, 100 °C, 24 h) of cecropin A was analysed at the Institute of Biochemistry, Uppsala, Sweden. The results are expressed as mol amino acid per mol protein. These values were calculated with a program that minimizes the deviations from integral numbers.

* Including the corresponding amide.

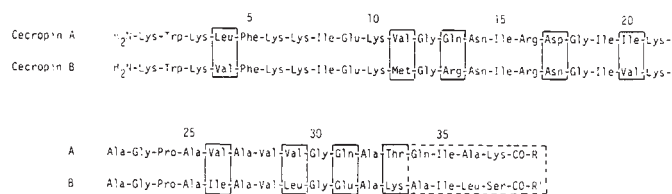


Fig. 1 The tentative amino acid sequence for cecropins A and B. Replacements of amino acids within the first 33 residues are boxed by solid lines. The C-terminal part boxed by a broken line is tentative. R and R' are unidentified blocking groups. The purified proteins used in sequence determination were obtained by collecting immune haemolymph 7 days after vaccination with live *Enterobacter cloacae* as previously described⁷. The haemolymph was applied to a Sephadex G-100 column equilibrated with 0.15 M ammonium acetate buffer, pH 5, containing phenylthiourea. The cecropins eluted together with the lysozyme. This material was pooled and further purified by ion-exchange chromatography⁷.

Both forms of cecropin contain one residue each of Trp and Phe. On this basis the theoretical absorbance (A_{280} , 1 mg ml^{-1}) was calculated to be 1.39 and 1.38 for cecropins A and B, respectively, which agrees well with the respective values of 1.36 and 1.31 determined experimentally, based on the protein concentration obtained from amino acid analysis. This means that an earlier value for cecropin A based on Lowry determination⁷ must now be disregarded.

The molecular weights (MWs) calculated from the sequence data (omitting the blocking groups) are 4,005 and 4,036 for cecropins A and B, respectively. Equilibrium centrifugation of cecropin A in 0.19 M ammonium formate, pH 6.5, gave a MW of $4,200 \pm 300$, which indicates a monomeric form. Furthermore, dimethylsuberimidate did not produce any cross-linking between cecropin molecules in conditions which showed melittin to be a tetramer. In contrast, gel filtration of cecropin A on a calibrated column of Sephacryl S-200 gave a MW of $\sim 14,000$. On Sephadex G-100, cecropins A and B eluted together with the lysozyme for which the MW was found to be 15,000 (ref. 7). Thus we cannot determine whether cecropin A is always a monomer.

The following points emerge from the structure of the cecropins. First, the sequences obtained for cecropins A and B show a strong homology; they share 25 amino acid residues and differ in only 12. This suggests that both originate from a gene duplication which may stretch only to residue 32. Assuming that the genetic code is universal, 7 out of 8 amino acid replacements in the first 32 residues can be explained by single shifts in the first base, while the remaining one, a Gln-Arg shift requires a single shift in the second base. More complicated genetic events are needed to explain the amino acid differences in the C-terminal part.

Second, the cecropin sequences have been compared with various other known protein sequences and the greatest

homology was found in residues 3–7 in cecropin A (Lys-Leu-Phe-Lys-Lys)—these are identical to residues 40–44 in the major coat protein of the filamentous phages fd and fl (ref. 11). Homology was also found in residues 25–28 (Ala-Val-Ala-Val) which are also present in the signal peptide of the λ receptor protein¹². It is possible that these two sequences could be involved in the membrane-crossing activities of the respective proteins.

Third, the amino acid composition and size of the cecropins are different from those of the bactericidal and basic proteins in polymorphonuclear leukocytes¹³ and from the antibacterial factors in seminal fluid¹⁴, saliva¹⁵ and milk¹⁶. Thus we conclude that cecropins A and B differ from other known antibacterial proteins. Similar compounds have been found in many other Lepidoptera¹⁷ and we suggest that the name cecropin should be applied to all antibacterial proteins which show a substantial homology with cecropins A and B.

We have previously shown that cecropins kill and lyse *Escherichia coli* and some other bacteria⁷. Using other methods we have now compared the biological activity of the cecropins with the bee venom toxin melittin (for a review see ref. 9). The nature of these two molecules is essentially similar; they are both small polypeptides with one part of the chain being strongly basic and one part predominantly hydrophobic. However, the C-terminal in melittin is basic, whereas it is the N-terminal that is basic in the cecropins.

The antibacterial effects of haemolymph, lysozyme, melittin and cecropins A and B were tested on six different bacteria (see Table 2). Activity against all organisms tested was induced in the haemolymph by vaccination with *Enterobacter cloacae*. Both cecropins had similar antibacterial spectra, although the B-form was significantly more active against *Bacillus subtilis*. Melittin showed no antibacterial activity against *Serratia marcescens* and *Xenorhabdus nematophilus* and for the remaining bacteria it had a lower activity than the cecropins. Differences in activity against the various organisms in Table 2 must be interpreted with care as some of the bacteria can produce proteolytic exoenzymes.

Figure 2 shows the bactericidal and lytic activities of cecropin A and melittin on *E. coli* D31 and on Chang liver cells. Melittin has a high lytic activity against both *E. coli* and the Chang liver cells whereas cecropin A was specific for the bacterial cells. Even at concentrations as high as $20 \mu\text{M}$ there was no detectable effect on the Chang liver cells, while a 300 times lower concentration gave 50% lysis of *E. coli*. The specificity of the cecropins agrees with earlier findings³ that immune haemolymph has no effect on sheep erythrocytes or insect tissue culture cells (unpublished data).

We have drawn two conclusions from our experiments. First, that the cecropins are the main antibacterial components so far identified in immune haemolymph which act against potential insect pathogens like *S. marcescens*, *Pseudomonas aeruginosa*

Table 2 Antibacterial activities of lysozyme, cecropins A and B and melittin

Organism	Strain	Diameter of inhibition zone (mm)						
		NH	IH	Lysozyme	Cecropin A	Cecropin B	Melittin	Blank
<i>E. coli</i> K12	D31	2.7*	13.2	2.7*	14.0	12.5	7.6	2.7*
<i>S. marcescens</i>	Db1108	2.7*	6.2	2.7*	7.6	7.2	2.7*	2.7*
<i>P. aeruginosa</i>	OT97	2.7*	3.9	2.7*	8.6	9.7	4.0	2.7*
<i>X. nematophilus</i>	Xn21	2.7*	5.0	2.7*	11.0	10.0	2.7*	2.7*
<i>Bacillus megaterium</i>	Bm11	2.7*	9.5	7.7	10.0	10.0	8.0	2.7*
<i>E. coli</i> K12	D31		13.0		18.0	17.0		
<i>B. subtilis</i>	Bs11		5.2		2.7*	7.2		

Antibacterial activity was recorded as inhibition zones on thin agar plates of a rich medium plus $\sim 8 \times 10^4$ viable cells per 6 ml. As all strains are streptomycin resistant, all the plates contained streptomycin ($100 \mu\text{g ml}^{-1}$). The concentration of cecropins A and B and melittin was 0.13 mM in all cases except the last two, where the cecropin concentration was 0.55 mM . The concentration of the cecropia lysozyme was 0.03 mM . The volume applied to each well was $3 \mu\text{l}$. Normal haemolymph (NH), immune haemolymph (IH) and a blank with 0.4 M ammonium formate buffer were used as references. IH, lysozyme and the cecropins were prepared as described in Fig. 1 legend. The diameter of the inhibition zones is proportional to the log of the concentration of inhibiting agent¹⁷. Well diameter 2.7 mm .

* No inhibition was detected.

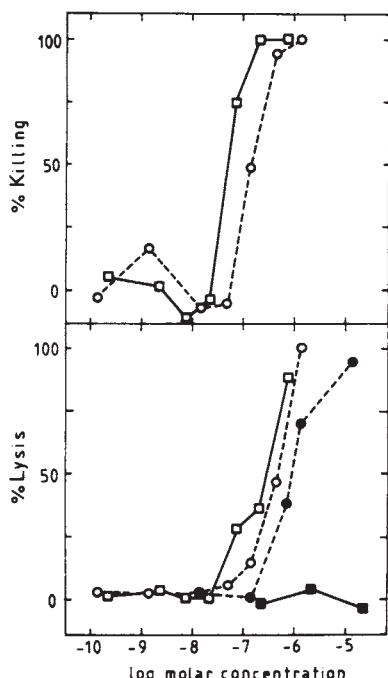


Fig. 2 Comparison of the action of cecropin A (squares) with melittin (circles) on *E. coli* D31 (open symbols) and Chang liver cells (closed symbols). *E. coli* (10^8 viable cells per ml) was incubated with the bacteriolytic agent at 37 °C in 0.1 M phosphate buffer, pH 6.4. After 30 min the tubes were put on ice and samples were withdrawn and diluted for viable count (given as per cent killing in the upper part of the figure). The remaining incubation mixture was centrifuged, and A_{260} of the supernatant was measured. Lysis is given as per cent of maximum release, corrected for the spontaneous release in buffer alone. Growth of strain D31 and other details were as before⁷. Chang liver cells were cultured in Parker 199 medium containing 5% fetal calf serum (FCS). The cells were washed and labelled with $\text{Na}_2^{51}\text{CrO}_4$ (0.1 mCi ml^{-1}) at 37 °C for 1 h in a Tris-Hanks buffered solution containing 3% FCS. The monolayer was then trypsinated and washed (R. Alsheikhly, in preparation). The lytic agent was added to 10^5 cells in 0.1 ml Tris-Hanks medium. The percentage of ^{51}Cr in the supernatant was measured after a 2-h incubation at 37 °C.

and *X. nematophilus*. A comparative study¹⁷ which included the greater wax moth, *Galleria mellonella*, and seven other Lepidoptera showed that cecropin-like substances are widely distributed in this order. As lysozyme is bactericidal for only a limited number of Gram-positive bacteria we believe that its main function is to remove the debris remaining after lysis of bacteria by cecropin.

Second, the cecropins must have a structure that prevents attack on the insect itself. This structural information is presumably absent in melittin; bees avoid self destruction by synthesizing an inactive precursor which is activated by a sequential liberation of dipeptides¹⁸. Future studies of the sequences of other cecropins should yield more information about the recognition of self in insects as well as an understanding of the bactericidal mechanism of the cecropins.

We thank Dr. A. R. Alsheikhly for help in the experiments with Chang cells. This work was supported by grants from the Swedish Natural Science Research Council (BU2453 to H.G.B.) and from the Swedish MRC (13P-5057 and 13X-3556 to H.B.).

Received 2 January; accepted 6 May 1981.

1. Boman, H. G. in *Microbial Control of Insects, Mites and Plant Diseases*, 1970-1980, 769-784 (Academic, New York, 1981).
2. Boman, H. G., Boman, A. & Pilon, A. *Insect Biochem.* **11**, 33-42 (1981).
3. Rasmuson, T. & Boman, H. G. in *Developmental Immunobiology*, 83-90 (Elsevier, Amsterdam, 1977).
4. Powning, R. F. & Davidson, W. J. *Comp. Biochem. Physiol.* **B55**, 221-228 (1976).
5. Jollès, J., Schoentgen, F., Croizier, G., Croizier, L. & Jollès, P. *J. molec. Evol.* **14**, 267-271 (1979).
6. Jarosz J. *Biol. Zbl.* **98**, 459-471 (1979).

7. Hultmark, D., Steiner, H., Rasmuson, T. & Boman, H. G. *Eur. J. Biochem.* **106**, 7-16 (1980).
8. Chadwick, J. M. & Aston, W. P. in *Animal Models of Comparative and Developmental Aspects of Immunity and Disease*, 1-14 (Pergamon, New York, 1978).
9. Mackler, B. F. & Kreil, G. *Inflammation* **2**, 55-65 (1977).
10. Inglis, A., McKern, N., Roxburgh, C. & Strike, P. in *Methods in Peptide and Protein Sequence Analysis*, 329-343 (Elsevier, Amsterdam, 1980).
11. Webster, R. E. & Cashman, J. S. in *The Single-Stranded DNA Phages*, 557-569 (Cold Spring Harbor Laboratory, New York, 1978).
12. Hedgpeth, J., Clément, J. M., Marchal, C., Perrin, D. & Hofnung, M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2621-2625 (1980).
13. Spitznagel, J. K. in *Salvia and Dental Caries*, 391-411 (Information Retrieval, New York, 1978).
14. Reddy, E. S. P. & Bhargava, P. M. *Nature* **279**, 725-728 (1979).
15. Mandel, I. D. in *Salvia and Dental Caries*, 473-491 (Information Retrieval, New York, 1978).
16. Harmon, R. J. in *Salvia and Dental Caries*, 413-427 (Information Retrieval, New York, 1978).
17. Hoffmann, D., Hultmark, D. & Boman, H. G. *Insect Biochem.* (in the press).
18. Kreil, G., Haiml, L. & Suchanek, G. *Eur. J. Biochem.* **111**, 49-58 (1980).

Is membrane expansion relevant to anaesthesia?

N. P. Franks

Biophysics Section, Department of Physics, Imperial College of Science and Technology, Prince Consort Road, London SW7 2AZ, UK

W. R. Lieb

Department of Biophysics, King's College, 26-29 Drury Lane, London WC2B 5RL, UK

General anaesthesia can be induced by a wide variety of structurally dissimilar molecules. Consequently, the mechanism must involve some rather nonspecific interactions at the target site, generally held to be in nerve membranes. The primary site of action has been postulated to be either lipid or protein or both. Although recent work¹⁻⁴ has cast doubt on the lipid hypotheses, protein models still flourish. In particular, Seeman and his co-workers⁵⁻⁸ have shown that biological membranes expand when anaesthetic molecules are added, and that this expansion is far greater than that which occurs with lipid bilayers. It has been suggested that this difference is due to extensive conformational changes in the membrane proteins, and several mechanisms have been proposed^{6,9,10} to explain this large expansion of proteins. We now report the first direct measurements of the volumes occupied by general anaesthetic molecules in both biological membranes and lipid bilayers. We show that, in fact, biological membranes expand much less and lipid bilayers expand more than previously reported. The volume that a general anaesthetic molecule occupies is essentially the same in biological membranes, lipid bilayers and water. Our results lead us to question all generalized membrane expansion hypotheses for the mechanism of general anaesthesia, in favour of hypotheses which include more specialized target sites.

The approach we used consisted of determining membrane densities and membrane/buffer partition coefficients as a function of anaesthetic concentration. A straightforward and extremely accurate method of measuring solution densities, using a vibrating tube, has been developed by Kratky and co-workers¹¹. Commercially available instruments are capable of measuring densities to an accuracy of about $10^{-6} \text{ g cm}^{-3}$. By measuring the density ρ_{sus} of a membrane suspension and the density ρ_{buf} of the buffer alone, the density of the membrane can be calculated as

$$\rho_{\text{mem}} = \frac{\rho_{\text{sus}} \rho_{\text{buf}}}{\rho_{\text{buf}} - R(\rho_{\text{sus}} - \rho_{\text{buf}})} \quad (1)$$

where R is the ratio of buffer mass to membrane mass. (This equation can be derived from the simple relationship $v_{\text{mem}} + Rv_{\text{buf}} = (1 + R)v_{\text{sus}}$, where v stands for specific volume, the reciprocal of density.) Seeman⁸, using such an instrument, has

reported large changes in the density of biological membranes but only very small changes in the density of lipid bilayers due to the presence of ethanol. As some of his data have recently been questioned¹², we began by repeating these measurements.

Our results using ethanol are shown in Fig. 1. The curves in Fig. 1a show the progressive decrease in densities of both buffer and a suspension of human red blood cell membranes, as a

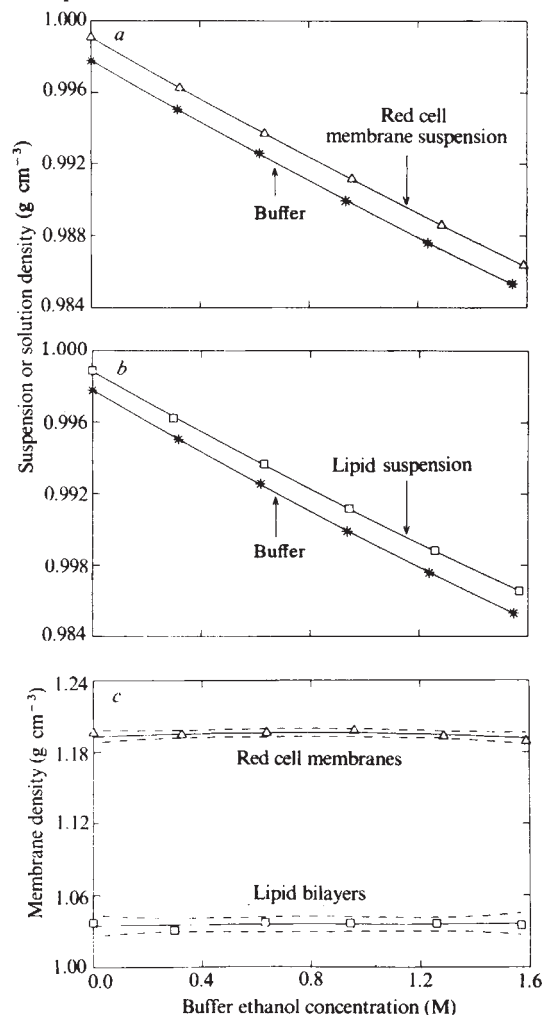


Fig. 1 Membrane density measurements with ethanol. *a*, Density of human red blood cell membrane suspension (Δ) and of buffer (*) as a function of the buffer ethanol concentration. *b*, Density of bovine spinal cord lipid suspension (□) and of buffer (*) versus buffer ethanol concentration. *c*, Density of red cell membranes (Δ) and spinal cord lipid membranes (□) versus buffer ethanol concentration, together with 95% confidence envelopes (dashed lines). Red cell membranes were prepared from transfusion blood (<1 week old) according to the method of Dodge *et al.*¹⁹. Total spinal cord lipids were especially prepared by Lipid Products Ltd and supplied as a solution in chloroform/methanol. Most of the solvent was removed using a rotary evaporator; remaining traces were removed by pumping under high vacuum for 6 h. The dry lipids were dispersed in buffer by vortex action and brief immersion in a sonicating bath. The buffer used in all our experiments was 5 mM potassium phosphate (pH 8). Precise values for ethanol concentrations were obtained using $c = c_{stk}(M_{stk}/M)(\rho/\rho_{stk})$, where c , M and ρ were the ethanol concentration, the mass and the density, respectively, of the final solution or suspension, and c_{stk} , M_{stk} and ρ_{stk} were the corresponding quantities for stock solutions of ethanol in buffer, which were added to membrane suspensions or buffer. For the red cell suspension, a small correction (always <1%) in c was made for partitioning of ethanol into the membranes, using a value⁶ of 0.14 moles ethanol per kg membrane/moles ethanol per l buffer. To determine membrane weights, dry weights of the membrane suspensions were measured gravimetrically and corrected for buffer salt weight. The value of R (=buffer weight/membrane weight) for the red cell membrane suspensions was 126.6 ± 2.5 , and 33.33 ± 0.39 for the spinal cord lipid suspension; the errors are for 95% confidence limits. All density measurements were performed using a precision density meter (Anton Paar, Model DMA 02C) at 25 °C. Temperature was maintained constant to within ± 0.01 °C. The curves in *a* and *b* were fitted to quadratics using the method of least squares. The curves and data points in *c* were calculated from the appropriate curves and data points of *a* and *b* using equation (1). The 95% confidence envelopes were calculated by least squares fitting of the data points to a quadratic.

function of the ethanol concentration in the buffer. Figure 1b gives the analogous results using bovine spinal cord lipids. Finally, Fig. 1c shows red cell membrane and spinal cord lipid densities, together with 95% confidence envelopes, as a function of the buffer ethanol concentration.

The most obvious feature of Fig. 1c is the clear difference in density between red cell and lipid membranes, due, of course, to the fact that red cell membranes contain about 50% protein. More important in the present context is the fact that, as regards the change in density with ethanol concentration, there is no significant difference between red cell membranes and spinal cord lipids. Indeed, there is no significant change in the density of either membrane over the range of ethanol concentrations we used. Our data are in direct disagreement with the experimental results reported by Seeman⁸, which showed a large (5%) change in red cell membrane density over the same range of ethanol concentrations.

It is important to appreciate that there is no simple relationship between changes in membrane density and changes in membrane volume. For example, if an anaesthetic has the same density in a membrane as the membrane itself, large expansions of membrane volume could occur without any change in density. The change in membrane volume can be calculated if one knows the partial molar volume, $\bar{V}_{an}$, of the anaesthetic as a function of its membrane concentration c_{mem} . $\bar{V}_{an}$, which has units of volume per mol, represents the change in membrane volume per mol of anaesthetic at a given anaesthetic concentration. The partial molar volume is directly related to changes in membrane density by the expression¹³

$$\bar{V}_{an} = \frac{M_{an} - (\partial \rho_{mem} / \partial c_{mem})}{\rho_{mem} - c_{mem}(\partial \rho_{mem} / \partial c_{mem})} \quad (2)$$

where M_{an} is the molecular weight of the anaesthetic, and ρ_{mem} is the membrane density at an anaesthetic molar concentration in the membrane of c_{mem} . If the membrane density is observed to change linearly with concentration, that is, if $\rho_{mem} = \rho_{mem}^0 + mc_{mem}$, then equation (2) reduces to

$$\bar{V}_{an} = (M_{an} - m) / \rho_{mem}^0 \quad (3)$$

where m is the slope of the curve of membrane density versus c_{mem} , and ρ_{mem}^0 is the membrane density in the absence of anaesthetic. Using this equation and the red cell membrane data from Fig. 1c, and taking a membrane/buffer partition coefficient⁶ of 0.14, $\bar{V}_{an}$ for ethanol lies between 20 and 65 cm³ mol⁻¹, with 95% confidence. Furthermore, the mean $\bar{V}_{an}$ for ethanol in the spinal cord lipids also falls within this range. Now, the partial molar volume of ethanol in buffer, calculated directly from the buffer density data of Fig. 1a,b, varies between 53.5 and 55.1 cm³ mol⁻¹ over the ethanol concentrations used. Thus, there seems to be no dramatic difference in the volume that an ethanol molecule occupies in red cell membranes, lipid bilayers and water.

However, ethanol is not the best choice of anaesthetic molecule if one is interested in measuring partial molar volumes in membranes. This is because, being relatively polar, ethanol has a very small membrane/buffer partition coefficient K . Consequently, only indirect and uncertain estimates of K exist or can be easily obtained. Furthermore, ethanol is not an anaesthetic which is used clinically. We therefore repeated our density measurements using the apolar anaesthetic halothane (Fig. 2) which is both widely used clinically and for which we were able to measure accurate membrane/buffer partition coefficients as a function of anaesthetic concentration. Note that the solution and suspension densities all increase with halothane concentration, whereas the opposite behaviour (decreasing density) was seen with ethanol concentrations (see Fig. 1). This, of course, reflects the fact that liquid halothane ($\rho = 1.87$ g cm⁻³) is much denser than membranes and buffer, whereas ethanol ($\rho = 0.79$ g cm⁻³) is less dense. Figure 2c shows the red cell membrane and spinal cord lipid densities (with 95% confidence envelopes), calculated using the data from Fig. 2a, b and equation (1).

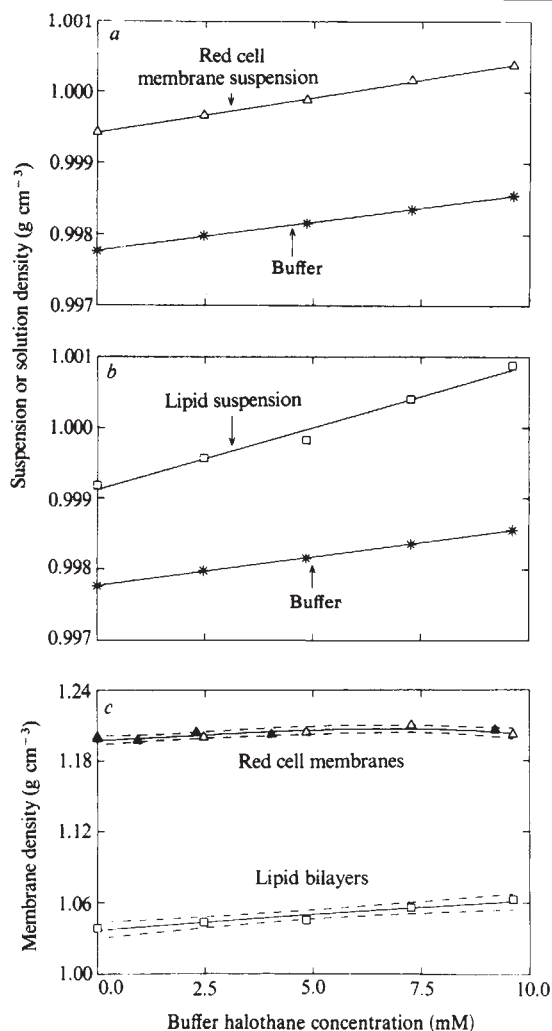


Fig. 2 Membrane density measurements with halothane. *a*, Density of human red blood cell membrane suspension (Δ) and of buffer ($*$) as a function of the buffer halothane concentration. *b*, Density of bovine spinal cord lipid suspension ($\square$) and of buffer ($*$) versus buffer halothane concentration. *c*, Density of red cell membranes (Δ and $\blacktriangle$) and spinal cord lipid membranes ($\square$) versus buffer halothane concentration, together with 95% confidence envelopes (dashed lines). The membrane suspensions were prepared and the density measurements performed as described in Fig. 1 legend. Known partial pressures³ of halothane in nitrogen were bubbled through the buffer and then passed over the membrane suspension, which was contained in a small (5 ml) round-bottomed flask. The suspension was stirred, using a glass-encapsulated magnetic 'flea'. This had the important technical advantage that the buffer concentrations of anaesthetic in equilibrium with the membranes was known directly and did not have to be corrected for partitioning into the membranes. The equilibrium concentration of halothane in the buffer and the aqueous phase of the membrane suspension was calculated² using an Ostwald solubility (water/gas partition) coefficient²⁰ of $\lambda = 1.5$ (at 25 °C). To avoid loss of halothane to air, the solutions were directly transferred to the density meter in a syringe containing no air pockets. The value of R ($\equiv$ buffer weight/membrane weight) was corrected for the differential partitioning of halothane between membranes and buffer using $R = R_0(1 + M_{an}c_{but}/\rho_{but}^0)/(1 + M_{an}Kc_{but}/\rho_{mem}^0)$, where R_0 is the value of R when $c_{but} = 0$. R_0 was 100.0 ± 1.9 (Δ), 74.3 ± 3.4 ($\blacktriangle$) and 26.55 ± 0.24 ($\square$); the errors are for 95% confidence limits. Partition coefficients K ($\equiv$ moles halothane per l membrane divided by moles halothane per l buffer) were determined using ^{14}C -labelled halothane (NEN) and equilibrium dialysis at 25 °C. The labelled halothane was repeatedly washed with buffer until its halothane/buffer partition coefficient was constant and consistent with its known solubility in water. For the red cell membrane determinations, SDS-polyacrylamide gels run before and after dialysis showed that no significant hydrolysis of membrane proteins had occurred. The measurements were performed as a function of aqueous concentration of halothane from zero to about 8 mM and fitted to least-squares polynomial curves of the statistically appropriate order²¹. The results were $K = 29.5 + 0.84c_{but}$ for spinal cord lipids and $K = 38.4 - 5.24c_{but} + 0.770c_{but}^2$ for red cell membranes, where c_{but} has units of moles halothane per l buffer. The density curves in *a* and *b* were fitted to straight lines using the method of least squares, and the curves and data points (Δ , $\square$) in *c* were calculated from the appropriate curves and data points of *a* and *b* using equation (1). The additional data points ($\blacktriangle$) in *c* are the results from a completely independent red cell membrane experiment and give a measure of the reproducibility of the data. For the red cell membranes two data points are superimposed at zero halothane concentration. The 95% confidence envelopes were calculated by least-squares fitting of the data points.

The data in Fig. 2c clearly show that there is a significant increase in the densities of both red cell membranes and lipid bilayers with increasing halothane concentration (except in the very highest concentration range for red cell membranes—see later). Accurate values of the partial molar volumes $\bar{V}_{an}$ can be obtained using these data if the concentrations of the anaesthetic in the membranes are known. To convert buffer concentrations to membrane concentrations, the partition coefficient K (mols halothane per l membrane divided by mols halothane per l buffer at equilibrium) had to be determined. We measured K as a function of halothane concentration for both red cell and lipid membranes, using equilibrium dialysis and ^{14}C -labelled halothane. Combining these results with the density measurements shown in Fig. 2c, we could calculate membrane density as a function of membrane halothane concentration (Fig. 3). The relationships between membrane density and membrane halothane concentration do not differ significantly from straight lines up to about 300 mM. Therefore, we could calculate $\bar{V}_{an}$ in this range using equation (3), for both red cell and lipid bilayer membranes. Using equation (3) with $M_{an} = 197.4 \text{ g mol}^{-1}$ and taking m and ρ_{mem}^0 as the slopes and intercepts of the least squares lines of Fig. 3, the values ($\pm 95\%$ confidence limits) are $\bar{V}_{an} = 134 \pm 14 \text{ cm}^3 \text{ mol}^{-1}$ for human red blood cell membranes and $\bar{V}_{an} = 124 \pm 24 \text{ cm}^3 \text{ mol}^{-1}$ for bovine spinal cord lipids. Thus, there is no significant difference between the partial molar volume of halothane in red cell and lipid membranes. It follows that, at equal membrane concentrations, halothane expands both membranes by the same amount. Furthermore, $\bar{V}_{an}$ in these membranes is about the same as it is in buffer ($116.7 \pm 3.2 \text{ cm}^3 \text{ mol}^{-1}$ at 95% confidence, calculated using the buffer data of Fig. 2).

Our results using halothane thus provide a firm quantitative basis for the conclusions which could be drawn from our experiments with ethanol. We find virtually no difference in the volume that a given anaesthetic molecule occupies in biological membranes, lipid bilayers or water. This was true for both halothane and ethanol, two very different general anaesthetics. We should point out that the range of anaesthetic concentrations used in our experiments included not only general anaesthetic but also nerve blocking levels.

Our measurements throw new light on membrane expansion theories of anaesthesia. These theories have the attraction of providing a conceptually simple framework for understanding the phenomenon of pressure reversal of general anaesthesia. The most popular of these, the critical volume hypothesis, states¹⁴ that "Anaesthesia occurs when the volume of a hydrophobic region is caused to expand beyond a certain critical amount by the absorption of molecules of an inert substance. If the volume of this hydrophobic region can be restored by changes of temperature or pressure, then the anaesthesia will be removed." At concentrations of halothane that produce general anaesthesia in man ($c_{but} = 0.27 \text{ mM}$), the expansion ($\bar{V}_{an}c_{mem}$) calculated from the data of Fig. 3 for both membranes and bilayers is only about 0.1%. This observed expansion is very small, in view of recent measurements¹⁵ which show that cholesterol-containing lipid bilayers expand by about 0.1% per °C. Thus, were the critical volume theory to be applied to whole membranes, it would predict that a rise in temperature of only a few degrees should induce general anaesthesia. This is not observed in cold-blooded animals. (A similar line of reasoning^{2-4,16,17} has been used to discredit both lipid phase transition models and bilayer fluidization hypotheses of anaesthesia.)

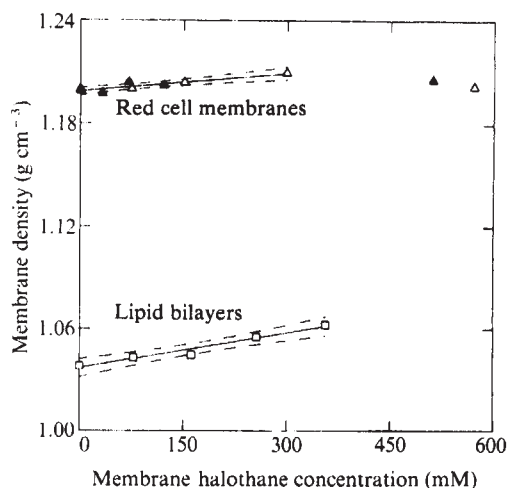


Fig. 3 Membrane density as a function of membrane halothane concentration and the determination of the partial molar volume $\bar{V}_{an}$ for human red cell membranes (upper line) and bovine spinal cord lipids (lower line). The slope and intercept of each least-squares straight line was used to calculate $\bar{V}_{an}$ with equation (3); the 95% confidence limits were calculated using these values and the variance-covariance matrix²². The red cell data points above 500 mM, where we suspect that some protein denaturation had occurred²³, were not included in the analysis. For red cell membranes, $\bar{V}_{an} = 134 \pm 14 \text{ cm}^3 \text{ mol}^{-1}$; for lipid bilayers, $\bar{V}_{an} = 124 \pm 24 \text{ cm}^3 \text{ mol}^{-1}$.

Furthermore, our observation that the volume occupied by a halothane molecule is essentially the same in membranes, water and liquid halothane shows that although membranes do indeed expand, the expansion is almost wholly accounted for by the volume of the halothane molecules themselves and does not, as previously thought⁸ for biological membranes, involve the creation of a large additional free volume which might then be compressed by the application of high pressures.

Overall, we conclude that the expansion of whole membranes or lipid bilayers is not relevant to the mechanism of anaesthesia. On the other hand, our results are consistent with the idea that general anaesthesia is caused by interactions with specialized regions of nerve membranes, which have properties quite different from those we have measured for whole membranes. Indeed, the partial molar volume of an anaesthetic molecule in these regions may well be much larger than in the rest of the membrane or in water; pressure reversal could then be simply a consequence of forcing the anaesthetic away from its target¹⁸. We have previously shown^{2,3}, using X-ray and neutron diffraction, that lipid bilayer structure is not significantly affected by general anaesthetics at up to 10 times clinical concentrations. Thus, it seems unlikely that general anaesthesia involves any sizeable structural change in the lipids of such specialized regions. This leaves us to suppose that it is the structure of particularly sensitive membrane proteins which are directly affected. For example, it could be that the target site is a multisubunit protein immersed in the membrane and that anaesthetics interfere with its function by interposing themselves between the subunits and disrupting the packing. The fact that general anaesthesia can be induced by a wide variety of structurally dissimilar molecules, whereas most drug/receptor interactions are rather specific, follows directly from such an idea. (A specific key is necessary to unlock a door, but anything can be used to prevent it from closing.)

We thank Dr E. G. Richards for stimulating discussions throughout this work, Dr W. B. Gratzer for help with the equilibrium dialysis measurements, Mr V. Ohanian for running the polyacrylamide gels, and Professor D. M. Blow for helpful comments on the manuscript. This research was supported by a project grant from the MRC and by a grant from the British Oxygen Corporation Limited.

5. Seeman, P., Kwant, W. O. & Sauks, T. *Biochim. biophys. Acta* **183**, 499–511 (1969).
6. Seeman, P. *Pharmac. Rev.* **24**, 583–655 (1972).
7. Seeman, P. & Roth, S. *Biochim. biophys. Acta* **255**, 171–177 (1972).
8. Seeman, P. *Experientia* **30**, 759–760 (1974).
9. Ueda, I., Kamaya, H. & Eyring, H. *Proc. natn. Acad. Sci. U.S.A.* **73**, 481–485 (1976).
10. Eyring, H., Woodbury, J. W. & D'Arrigo, J. S. *Anesthesiology* **38**, 415–424 (1973).
11. Kratky, O., Leopold, H. & Stabinger, H. *Meih. Enzym.* **27**, 98–110 (1973).
12. Richards, C. D. in *Topical Reviews in Anaesthesia* Vol. 1 (eds Norman, J. & Whitwam, J. G.) 1–84 (John Wright, Bristol, 1980).
13. Richards, E. G. *An Introduction to Physical Properties of Large Molecules in Solution* (Cambridge University Press, 1980).
14. Miller, K. W. & Smith, E. B. in *A Guide to Molecular Pharmacology-Toxicology* Vol. 1 (ed. Featherstone, R. M.) 427–475 (Dekker, New York, 1973).
15. Melchior, D. L., Scavitto, F. J. & Steim, J. M. *Biochemistry* **19**, 4828–4834 (1980).
16. Simon, S. A. & Bennett, P. B. in *Molecular Mechanisms of Anesthesia* Vol. 2 (ed. Fink, B. R.) 305–318 (Raven, New York, 1980).
17. Pang, K. Y., Braswell, L. M., Chang, L., Sommer, T. J. & Miller, K. W. *Molec. Pharmac.* **18**, 84–90 (1980).
18. Richards, C. D. *Int. Rev. Biochem.* **19**, 157–220 (1978).
19. Dodge, J. T., Mitchell, C. & Hanahan, D. J. *Archs Biochem. Biophys.* **100**, 119–130 (1963).
20. Regan, M. J. & Eger, E. I. *Anesthesiology* **28**, 689–700 (1967).
21. Snedecor, G. W. & Cochran, W. G. *Statistical Methods* 6th edn (Iowa State University Press, Ames, 1967).
22. Cleland, W. W. *Adv. Enzym.* **29**, 1–32 (1967).
23. Metcalfe, J. C., Seeman, P. & Burgen, A. S. V. *Molec. Pharmac.* **4**, 87–95 (1968).

Fast inward and outward current channels in a non-spiking neurone

Maurizio Miroli*

Békésy Laboratory of Neurobiology, 1993 East-West Road, University of Hawaii, Honolulu, Hawaii 96822, USA

Although the crustacean coxal receptors¹ are non-spiking^{2,3}, indirect pharmacological and electrophysiological evidence suggests that fast sodium channels may be present in their membrane^{4–6}. The properties of these channels are not known, but it has been suggested that they might be “incompletely differentiated”, perhaps lacking “appropriate gating mechanisms”⁵, and/or “too sparsely distributed”⁷. The former hypothesis is not supported by the results of voltage-clamping experiments done on dendritic segments isolated from these mechanoreceptors. Instead, the results reported here provide direct evidence for a voltage-dependent fast inward current, sensitive to tetrodotoxin (TTX) and requiring external sodium (but not calcium). This current is shunted by a transient fast outward current, also voltage dependent, and it is suggested that this shunting may account, at least in part, for the non-spiking behaviour of the coxal receptors.

The largest dendrite (T fibre¹) among those supplying the coxal muscle receptor organ in the crab *Portunus sanguinolentus* (Herbst) was ligated close to its sensory endings and to its entry into the thoracic ganglion, and then cut distal and proximal to the two ligatures. The input properties of the isolated dendrite were similar to those of intact coxal receptors³. In nine experiments the average and standard deviation of the resting potential were $63.8 \pm 5.5 \text{ mV}$ and those of the input resistance were $2.2 \pm 1.5 \text{ M}\Omega$. The current-voltage relationship was characterized by a pronounced outward-going rectification in the depolarizing quadrant. Depolarization by current pulses did not elicit overshooting action potentials (Fig. 1a).

Dendrites could be clamped at holding potentials close to or considerably more negative (up to -80 mV) than their resting potential for several hours without deterioration of their input properties. Depolarizing and hyperpolarizing clamping voltages of small amplitude resulted in symmetrical clamping currents, but at voltages more positive than about -50 mV , a transient inward current (I_N) became evident (Fig. 1b). Both amplitude and time to peak of I_N depended on the magnitude of the test clamping voltage: the amplitude first increased with larger values of the test voltage and then decreased (Fig. 1b), while the time to peak decreased from 3–5 ms to less than 1 ms with increasing intensity of the test voltages. I_N was also dependent on the holding voltage and was completely inactivated in dendrites held at voltages more positive than -55 mV .

I_N was sensitive to pharmacological agents known specifically to affect voltage-dependent sodium channels⁸. It was abolished

* Permanent address: Indiana University, School of Medicine, Medical Sciences Program, Physiology Section, Myers Hall, Bloomington, Indiana 47405, USA.

1. Buggs, J. M., Young, T. & Hsia, J. C. *Molec. Pharmac.* **12**, 127–135 (1976).
2. Franks, N. P. & Lieb, W. R. *Nature* **274**, 339–342 (1978).
3. Franks, N. P. & Lieb, W. R. *J. molec. Biol.* **133**, 469–500 (1979).
4. Richards, C. D. *et al. Nature* **276**, 775–779 (1978).

Received 16 March; accepted 30 April 1981.

when TTX in concentrations as low as 10^{-8} M was added to the saline (Fig. 1c, d), and when external sodium was substituted by choline, tetramethyl ammonium or tetraethyl ammonium. By contrast, I_N was not affected by agents known to interfere with calcium currents⁹. Thus, it was not increased when calcium was substituted by either barium or strontium, nor was it abolished or reduced when cadmium (10^{-4} M), manganese or cobalt (both 5×10^{-3} M) were added to the saline.

Even in TTX, clamping currents were not symmetrical in the hyperpolarizing and depolarizing range except, as noted, for small voltage steps. Instead, depolarizing voltages resulted in early fast outward current transients considerably larger than the capacitive transients obtained with hyperpolarizing tests of equal amplitude (Fig. 1c, d legend). For several criteria the fast outward current behaved in a manner similar to the fast inward current; thus, it could not be demonstrated in dendrites held at potentials more positive than about -45 mV, and its time to

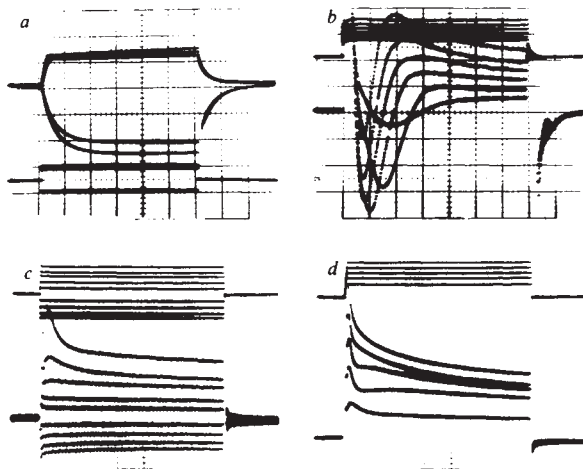


Fig. 1 Example of data obtained by current and voltage clamping in isolated dendritic segments. The preparations, 3–6 mm in length and 80–120 μ m in diameter, were pinned to the Sylgard floor of a chamber having an approximate volume of 0.016 ml, and perfused at the rate of about 0.5 ml min^{-1} with a saline solution of the following molar composition: Na, 0.47; K, 0.01; Ca, 0.026; Mg, 0.052; Cl, 0.513; SO_4 , 0.022, buffered to pH 7.6 with 0.005 mol of Na-HEPES. The voltage clamping circuit was modelled after that of Connor and Stevens¹⁰. The current injecting pipette was screened. Position and tip resistance of the two micropipettes used for clamping were critical in obtaining good results. Stable input properties and adequate clamping speed were obtained with pipettes having a resistance of 4–7 M Ω , inserted in the middle of the dendrites at about 50 μ m from each other. Fast clamping was not possible with micropipettes having higher resistance, while the use of micropipettes having lower resistance results in unstable input properties. In each record, voltage is recorded in the top trace and current in the lower one. Calibrations: one major grid division is equal to: a, 20 mV, 50 nA, 50 ms; b, 50 mV, 100 nA, 1 ms; c, 50 mV, 20 nA, 5 ms; d, 100 mV, 50 nA, 10 ms. a, Current clamping of a dendritic segment 5.8 mm long, whose membrane potential (V_m) was -65 mV and input resistance (R_i) 2.2 M Ω . Note lack of spiking activity and pronounced outward-going rectification above -45 mV. b, Voltage clamping of a segment 4.6 mm long (V_m -65 mV; R_i 1.75 M Ω). Inward current (I_N) corresponds to a downward deflection of the current trace. I_N reached a maximum value (360 nA) when the segment was depolarized to -35 mV from a holding potential of -75 mV. c, d, Outward currents recorded in saline containing tetrodotoxin (5×10^{-8} M) in a segment 3.6 mm long whose resting potential was -67 mV and input resistance 2.4 M Ω . The preparation was held at -80 mV except during testing. As shown in d, at clamping voltages more positive than -30 mV, the initial outward transient is followed by a second one characterised by a slower rate of decay. This second outward transient is not considered in the text because its time course is too slow to affect the inward current directly. Preliminary results indicate that the slower outward transient can be blocked by external tetraethyl ammonium (5×10^{-2} M) whereas the initial fast transient is not. Both outward currents are blocked by 4-aminopyridine (5×10^{-3} M).

peak decreased with increasing values of the test voltage, from 3–5 ms to <1 ms (Fig. 1c, d). Unlike the inward current, however, the early outward current transient increased monotonically in magnitude with larger values of the clamping voltage (Fig. 1c, d).

The shunting effect of the outward current transients on the inward current flowing through the TTX-sensitive channels can be evaluated from Fig. 2. The difference between clamping currents, recorded in the same preparation in control and in TTX saline, or net inward current, is substantially larger than I_N and has a null point at about $+50$ mV and thus at a voltage close to the reversal potential expected for a sodium current. I_N , on the other hand, has a null point at about -10 mV, close to the peak values of the partial action potentials which can occasionally be recorded from intact coxal receptors^{3,4}.

The net inward current density, calculated from data such as those presented in Fig. 2, on the assumption that the dendritic segments were properly space clamped, is about 20 $\mu\text{A cm}^{-2}$ and thus much lower than that measured in excitable membranes. It should be emphasised, however, that the validity of this assumption (as well as other assumptions involved in the

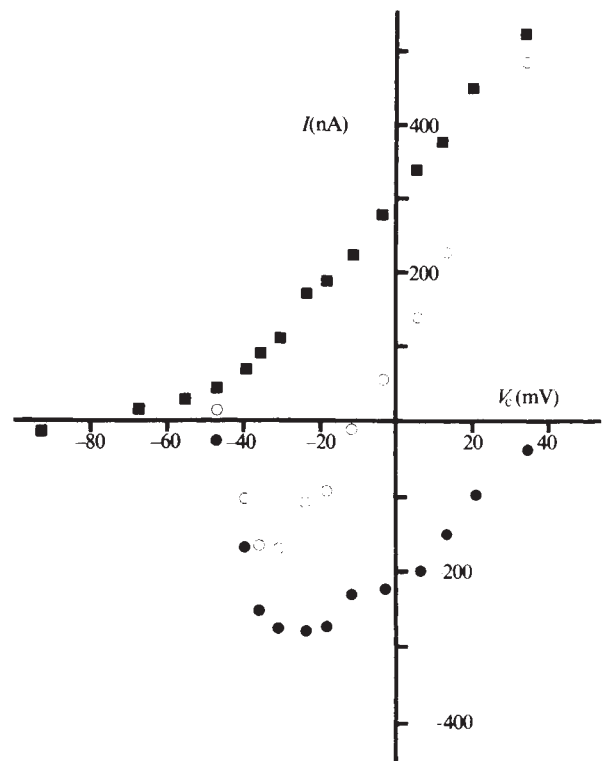


Fig. 2 Plot of clamping currents obtained from the same preparation in control saline ($\circ$) and in saline containing TTX ($\blacksquare$) in response to stimuli of equal amplitude (V_c). Currents were measured at times corresponding to peak amplitude of the inward current in control saline. Their difference, net peak inward current, is plotted ($\bullet$). Clamping was not sufficiently fast at voltages larger than about $+15$ mV, and therefore no reliable results were obtained close to the null potential of the inward current. By extrapolation, this potential can be seen to be about $+50$ mV. This dendritic segment was 6 mm long, and its surface was 1.5×10^{-2} cm^2 . As the length constant of the coxal receptor dendrites is at least 3 cm (refs 3, 7) the preparations should have been properly space clamped if the ligatures at their endings were to provide a complete seal¹⁸. On this assumption, peak net inward current density could be calculated to be, in this example, 19 $\mu\text{A cm}^{-2}$. Two other experiments yielded values of 18 and 48 $\mu\text{A cm}^{-2}$, and the null potentials were, respectively, $+25$ and $+43$ mV. Assuming further that the effective membrane capacitance of the dendrites was about 1 $\mu\text{F cm}^{-2}$, the current density thus calculated would be about 10 times lower than that required to charge the membrane to the null potential of the inward current. If the endings were leaky, the amount of membrane under clamp control would have been less, and hence current density would be underestimated by an unknown amount.

calculations) cannot be properly evaluated on the basis of the available data (see Fig. 2 legend). Thus, it seems reasonable to conclude that the inability of the coxal receptor dendrites to develop full action potentials does not depend on a lack of functional sodium channels but rather on the shunting of the sodium current by the transient opening of a fast outward conductance and, possibly, on an abnormally low density of the sodium channels themselves. These two factors are not mutually exclusive; on the contrary, both can be easily interpreted as an expression of the specialization of the coxal receptors for transmitting information with analogue signals^{2,3}.

Results consistent with those presented in Figs 1 and 2 were also obtained in the S dendrite of *Portunus*, as well as in the S and T dendrites of other swimming crabs. Moreover, a fast TTX-sensitive and sodium-dependent inward current and a fast initial outward transient could also be demonstrated when clamping intact *Portunus* coxal receptors. In the intact cells no fast clamping could be obtained at voltages more positive than about -20 mV, and therefore a satisfactory study of the fast currents has not been possible.

The coxal receptors are not unique in exhibiting transiently activated outward currents¹⁰⁻¹⁷. There is convincing evidence that currents of this type are important in regulating repetitive firing of normally spiking neurones^{10,14}. The early fast outward channels of the coxal receptors, characterised by activation and inactivation rates comparable with those of the inward channels, seem to be specialised for suppressing spiking. Whether this specialisation represents an extreme case of adaptation of membrane components also found in other cells, or whether it reflects unique ionic and pharmacological properties, remains to be determined.

Received 15 December 1980; accepted 12 May 1981.

1. Alexandrowicz, J. S. & Whitear, M. J. *J. mar. biol. Ass. U.K.* **36**, 603-628 (1957).
2. Ripley, S. H., Bush, B. M. H. & Roberts, A. *Nature* **218**, 1170-1171 (1968).
3. Mirolli, M. *J. exp. Biol.* **78**, 1-27 (1979).
4. Roberts, A. & Bush, B. M. H. *J. exp. Biol.* **54**, 515-524 (1971).
5. Lowe, D. A., Bush, B. M. H. & Ripley, S. H. *Nature* **274**, 289-290 (1978).
6. Bush, B. M. H., DiCaprio, R. A. & Taylor, P. S. *J. Physiol., Lond.* **303**, 20-21P (1980).
7. Bush, B. M. H. in *Neurones Without Impulses* (eds Roberts, A. & Bush, B. M. H.) 147-176 (Cambridge University Press, 1981).
8. Narahashi, T. *Physiol. Rev.* **54**, 813-889 (1974).
9. Akaïke, N., Lee, K. S. & Brown, A. M. *J. gen. Physiol.* **71**, 509-531 (1978).
10. Connor, J. A. & Stevens, C. F. *J. Physiol., Lond.* **213**, 1-19, 21-30 (1971).
11. Neher, E. *J. gen. Physiol.* **58**, 36-53 (1971).
12. Nakajima, S. & Onodera, S. *J. Physiol., Lond.* **200**, 161-185 (1969).
13. Thompson, S. H. *J. Physiol., Lond.* **265**, 465-488 (1977).
14. Aldrich, R. W., Getting, P. A. & Thompson, S. H. *J. Physiol., Lond.* **291**, 507-530 (1979).
15. Nakajima, S. & Kusano, K. *J. gen. Physiol.* **49**, 613-628 (1966).
16. Frankenhauser, B. & Walman, B. *J. Physiol., Lond.* **148**, 677-682 (1959).
17. Miyazaki, S. I., Ohmori, H. & Sasaki, S. *J. Physiol., Lond.* **246**, 55-78 (1975).
18. Norman, R. S. *Biophys. J.* **12**, 22-45 (1972).

Inhibition of smooth muscle tension by cyclic AMP-dependent protein kinase

W. Glenn L. Kerrick* & Phyllis E. Hoar

Department of Physiology and Biophysics, SJ-40, University of Washington, Seattle, Washington 98195, USA

β -adrenergic relaxation of smooth muscle by catecholamines has been associated with elevated levels of cyclic AMP^{1,2}. The question arises whether subsequent activation of cyclic AMP-dependent protein kinase^{3,4} has a role in the regulation of smooth muscle contraction. There is substantial evidence that a Ca^{2+} -activated myosin light chain kinase/phosphatase system regulates smooth muscle contraction⁵⁻¹⁵, and Adelstein *et al.*^{16,17} have shown that the catalytic subunit of cyclic AMP-dependent protein kinase^{3,4} plays a part in this regulation, by phosphorylation of the high molecular weight subunit of the light chain kinase, which results in a decrease in the activity of the kinase. Here we have shown for the first time that the catalytic subunit of the protein kinase inhibits Ca^{2+} -activated tension in skinned smooth muscle fibre preparations.

Figure 1a shows that after a submaximal ($p\text{Ca} = 5.0$) and a maximal ($p\text{Ca} = 3.6$) contraction, addition of 10^{-6} M catalytic subunit of cyclic AMP-dependent protein kinase results in a slow inactivation of the fibres. After transfer of the fibre bundle from this high $[\text{Ca}^{2+}]$ catalytic subunit solution to a relaxing solution ($p\text{Ca} = 8.0$) the fibres relaxed. When they were reimmersed in the same solution ($p\text{Ca} = 5.0$) in which they initially gave ~90% maximum tension, they generated only 14% of the subsequent tension in $p\text{Ca} 3.6$ solution. This shows that pretreatment with the catalytic subunit of protein kinase resulted in a decrease in the Ca^{2+} -sensitive threshold for contraction of the muscle fibre preparation. Transfer of the fibre preparation from high $[\text{Ca}^{2+}]$ solution ($p\text{Ca} = 3.6$) to an identical solution containing 5 μM calmodulin resulted in a further increase in tension.

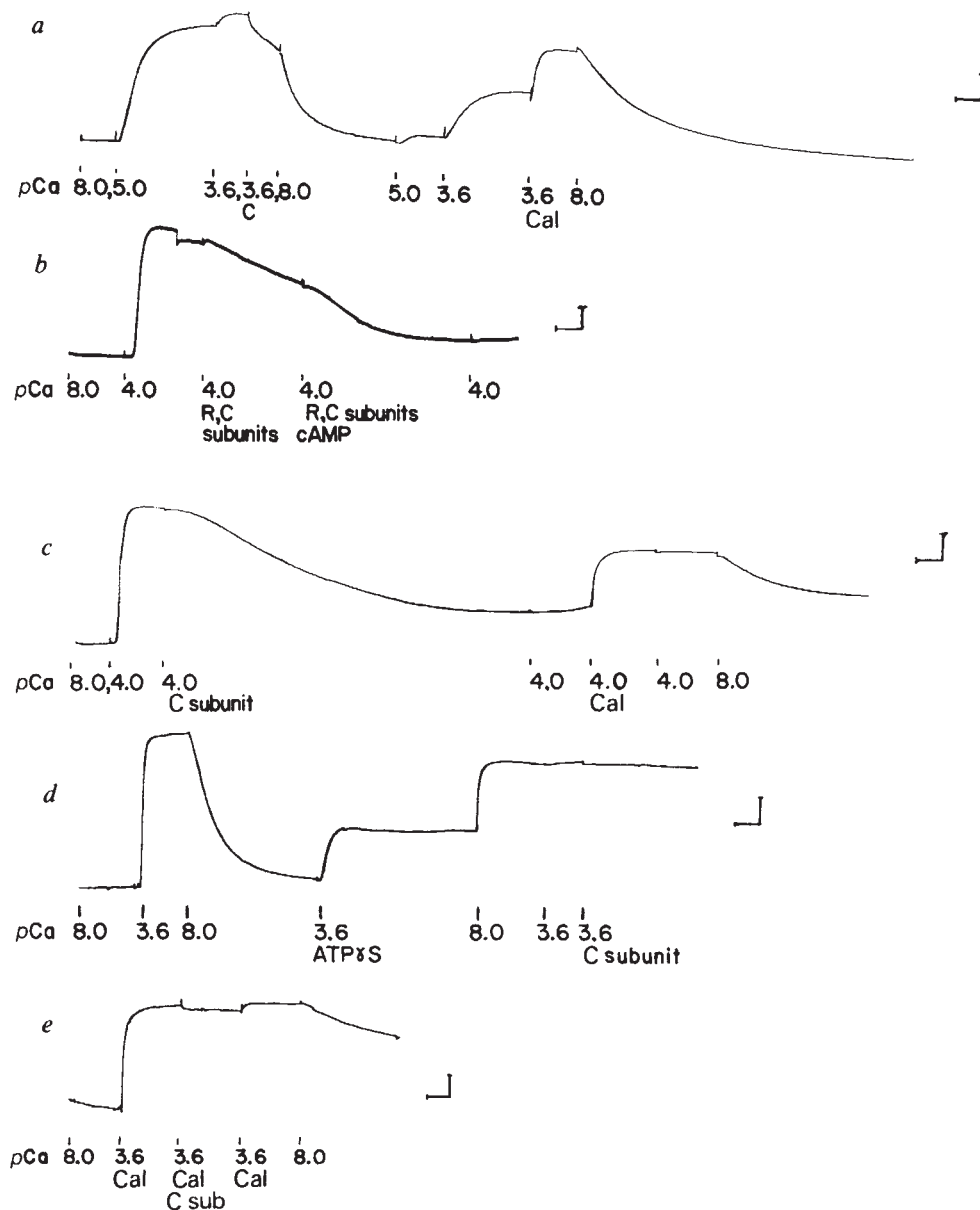
Figure 1b shows that after a maximal Ca^{2+} -activated contraction, addition of both the regulatory and catalytic subunits of protein kinase results in a slow inactivation of the fibres and that addition of cyclic AMP (10^{-4} M) produces faster relaxation. Cyclic AMP is known to activate protein kinase by dissociation of the catalytic subunit from the regulatory subunit⁴; the free catalytic subunit is then able to inhibit the fibres. Initial slow inactivation of the fibre bundle is presumably due to either the production by the fibres of small amounts of cyclic AMP from ATP, or incomplete association of mixed catalytic and regulatory subunits, or both. A higher than physiological concentration of cyclic AMP was used to avoid diffusional problems associated with cyclic AMP binding sites in the fibres. After maximal activation of the fibre bundle ($p\text{Ca} = 4.0$, Fig. 1c) and inhibition by the catalytic subunit, immersion of the fibres in a solution containing high $[\text{Ca}^{2+}]$ but no catalytic subunit resulted in little detectable activation of the fibres over the time period observed. However, in the presence of high $[\text{Ca}^{2+}]$ plus calmodulin, the fibres contracted immediately, giving partial recovery.

Figure 1d shows that after a skinned fibre bundle had been pretreated with high $[\text{Ca}^{2+}]$ and adenosine 5'-O-(3-thiotriphosphate) (ATP γS) to irreversibly thiophosphorylate the myosin LC₂₀ light chains, the fibre bundle became fully activated when immersed in a solution containing ATP and no calcium ($p\text{Ca} = 8.0$) and showed no further increase in tension in high $[\text{Ca}^{2+}]$ solution ($p\text{Ca} = 3.6$). Following this irreversible thiophosphorylation and tension activation, addition of the catalytic subunit was unable to inactivate the fibre bundle. Thus when the light chains are irreversibly thiophosphorylated¹⁰⁻¹³, the catalytic subunit of protein kinase can no longer inhibit the contraction. In the presence of 5 μM calmodulin (Fig. 1e), 1 μM catalytic subunit produced very little change in Ca^{2+} -activated tension generation, in contrast to its action in the absence of calmodulin. The results shown in Fig. 1 are typical of results obtained with either the skeletal or cardiac catalytic subunit of cyclic AMP-dependent protein kinase.

Our data thus agree with the light chain kinase/phosphatase mechanism for Ca^{2+} activation of smooth muscle and the proposed inhibition of this mechanism by the catalytic subunit of the cyclic AMP-dependent protein kinase¹⁶⁻¹⁸. The model proposed by Adelstein *et al.*^{16,17}, in which the phosphorylation of the high molecular weight subunit of light chain kinase by the catalytic subunit of the cyclic AMP-dependent protein kinase decreases the ability of the Ca^{2+} -binding subunit of calmodulin to interact with it, predicts that tension would be inhibited by this catalytic subunit and the inhibited fibres would appear less sensitive to calcium (Fig. 1a). Once the fibres have been inhibited partially (Fig. 1a) or completely (Fig. 1c), addition of calmodulin should increase the Ca^{2+} sensitivity of the muscle according to the law of mass action. Also, once the LC₂₀ light chains have been irreversibly thiophosphorylated by the use of the ATP analogue ATP γS ¹⁰⁻¹³, inhibition of the light chain kinase by the catalytic subunit should have no effect on tension development (Fig. 1d). As the only site of irreversible thiophosphorylation is myosin LC₂₀ (ref. 11), the site of action of cyclic AMP-dependent protein kinase cannot be the thin

* Present address: Department of Physiology and Biophysics, University of Miami, PO Box 016430, Miami, Florida 33101, USA.

Fig. 1 Tension-time records of chicken gizzard skinned fibres. All solutions contained ATP unless indicated by ATP γ S. Additions to the solutions are indicated below the pCa value and refer only to that solution, not to subsequent ones. *a*, Effect of 1 μ M cardiac catalytic (C) subunit of cyclic AMP-dependent protein kinase on submaximal Ca $^{2+}$ -activated tension. Recovery of maximum tension with 5 μ M calmodulin (Cal) is shown. Vertical calibration bar, 8.3 mg. *b*, Inhibition of Ca $^{2+}$ -activated tension by 1 μ M catalytic (C) subunit plus 2 μ M cyclic AMP-free regulatory (R) subunit of skeletal cyclic AMP-dependent protein kinase, followed by addition of cyclic AMP (cAMP, 10^{-4} M). Vertical calibration bar, 1.9 mg. The mechanical artefact seen during the initial pCa = 4.0 concentration was due to a small knock to the table. *c*, Complete inhibition of Ca $^{2+}$ -activated tension by 1 μ M skeletal catalytic (C) subunit of cyclic AMP-dependent protein kinase followed by partial recovery with 5 μ M calmodulin (Cal). Vertical calibration bar, 11.1 mg. *d*, Lack of tension response to 1 μ M catalytic (C) subunit of cardiac cyclic AMP-dependent protein kinase after pretreatment (irreversible thiophosphorylation of the 20,000-MW myosin light chains) of the fibre bundle with high [Ca $^{2+}$] (pCa = 3.6) and ATP γ S. Vertical calibration bar, 10.0 mg. *e*, Lack of inhibition by 1 μ M cardiac catalytic subunit (C sub) in the presence of 5 μ M calmodulin (Cal). Vertical calibration bar, 6.1 mg. The horizontal calibration bar represents 5 min in each case. Sections of muscle (~ 1 –2 mm 3) were cut from fresh chicken gizzards and lightly homogenized by a method described elsewhere for striated muscle 20 and previously used in this laboratory for gizzard skinned fibre preparations 11 . Small bundles of fibres were visualized by background light using a binocular microscope and the two ends of a fibre bundle were clamped in small stainless steel forceps of a tension transducer apparatus similar to that of Hellam and Podolsky 21 . The fibres were activated and relaxed by transferring them to different test bathing solutions containing various amounts of Ca $^{2+}$ and protein concentrations. The solutions contained 70 mM K $^{+}$ + Na $^{+}$, 1 mM Mg $^{2+}$, 2 mM Mg-ATP $^{2-}$, 15 mM creatine phosphate, 15 units of creatine phosphokinase (CPK) per ml solution, 82–84 mM imidazole, 7 mM EGTA total, varying amounts of Ca $^{2+}$ (pCa = 8–3.6) and propionate as the major anion. The pH was kept constant at 7.00 (± 0.02) and the ionic strength kept at 0.15 by varying the amount of imidazole propionate. However, the solution in which ATP γ S was substituted for ATP contained 85 mM K $^{+}$ + Na $^{+}$ and no creatine phosphate/CPK regenerating system, but was otherwise identical to the solutions described above. The total concentrations of magnesium, potassium and calcium propionate, disodium ATP, disodium creatine phosphate and imidazole required were determined by a computer program using binding constants obtained from the literature 22 .



filament, proposed by Nonomura and Ebashi 19 to regulate smooth muscle by a leiotoxin system. Furthermore, the presence of calmodulin prevents the inhibition of tension by the catalytic subunit (Fig. 1e). If phosphorylation of the 125,000-molecular weight (MW) light chain kinase decreases its interaction with calmodulin 17 , the presence of added calmodulin should overcome this by the laws of mass action. We will present elsewhere a detailed study of the relationship between increased phosphorylation of myosin light chain kinase, decreased phosphorylation of LC $_{20}$ light chain and inhibition of tension generation.

We have treated two other smooth muscle types, skinned using Triton X-100, with the catalytic subunit of protein kinase and obtained the same results as for chicken gizzard. These muscles were rabbit pulmonary artery and the aorta from the Pacific Coast dogfish which evolved 400 Myr ago. Therefore, the mechanism for this inhibition of tension in smooth muscle by cyclic AMP-dependent protein kinase seems to have evolved before much of the animal kingdom as we know it today.

These data show for the first time that the catalytic subunit of protein kinase can inhibit the physiological variable, Ca $^{2+}$ -

activated tension, in skinned smooth muscle fibre preparations. Pretreatment of the smooth muscle fibres with this catalytic subunit causes a decrease in the apparent Ca $^{2+}$ sensitivity of the muscles, an effect which can be reversed by the addition of calmodulin. In the presence of both the catalytic and regulatory subunits of protein kinase, addition of cyclic AMP increased the rate of inhibition of tension. These results agree with the mechanism described elsewhere 16,17 , which involves this protein kinase system in the β -adrenergic relaxation of smooth muscle, but do not support the thin filament control mechanism proposed by Nonomura and Ebashi 19 .

We thank L. Bolles and R. Coby for technical assistance, Dr J. Beavo for gifts of cyclic AMP-free regulatory subunit of the peak-1 bovine skeletal muscle cyclic AMP-dependent protein kinase and bovine skeletal catalytic subunit; Drs E. H. Fischer and E. Krebs for bovine cardiac catalytic subunit of cyclic AMP-dependent protein kinase, and Drs E. H. Fischer and D. Malencik for gifts of bovine brain calmodulin. This research was supported by grants from the Muscular Dystrophy Association, American Heart Association and NIH. P.E.H. was supported by a fellowship from the Muscular Dystrophy Asso-

ciation. The chicken gizzard muscles were supplied by Acme Poultry Co.

Received 30 October 1980; accepted 12 May 1981.

- Robison, G. A., Butcher, R. W. & Sutherland, E. W. *Cyclic AMP* 204–210 (Academic, New York, 1971).
- Kramer, G. L. & Hardman, J. G. *Handbook of Physiology* Sect. 2, Vol. 2 (eds Bohr, D. F., Somlyo, A. P. & Sparks, H. V.) 179–199 (American Physiological Society, Bethesda, 1980).
- Greengard, P. *Science* **199**, 146–152 (1978).
- Krebs, E. G. & Beavo, J. A. *Rev. Biochem.* **48**, 923–959 (1979).
- Aksoy, M. O., Williams, D., Sharkey, E. M. & Hartshorne, D. J. *Biochem. biophys. Res. Commun.* **69**, 35–41 (1976).
- Chacko, S., Conti, M. A. & Adelstein, R. S. *Proc. natn. Acad. Sci. U.S.A.* **74**, 129–133 (1977).
- Sobieszek, A. *Eur. J. Biochem.* **73**, 477–483 (1977).
- DiSalvo, J., Gruenstein, E. & Silver, P. *Proc. Soc. exp. Biol. Med.* **158**, 410–414 (1978).
- Sherry, J. M. F., Górecka, A., Aksoy, M. O., Dabrowska, R. & Hartshorne, D. J. *Biochemistry* **17**, 4411–4418 (1978).
- Cassidy, P., Hoar, P. E. & Kerrick, W. G. L. *J. biol. Chem.* **254**, 11148–11153 (1979).
- Hoar, P. E., Kerrick, W. G. L. & Cassidy, P. S. *Science* **204**, 503–506 (1979).
- Cassidy, P., Hoar, P. E. & Kerrick, W. G. L. *Pflügers Arch. ges. Physiol.* **387**, 115–120 (1980).
- Kerrick, W. G. L., Hoar, P. E. & Cassidy, P. S. *Fedn Proc.* **39**, 1558–1563 (1980).
- Kerrick, W. G. L., Hoar, P. E., Cassidy, P. S. & Malencik, D. A. in *Regulation of Muscle Contraction: Excitation-Contraction Coupling* (eds Grinnell, A. D. & Brazier, M. A. B.) (Academic, New York, in the press).
- Kerrick, W. G. L. in *Calcium and Cell Function* Vol. 2 (ed. Cheung, W. Y.) (Academic, New York, in the press).
- Adelstein, R. S., Conti, M. A., Hathaway, D. R. & Klee, C. B. *J. biol. Chem.* **253**, 8347–8350 (1978).
- Conti, M. A. & Adelstein, R. S. *Fedn Proc.* **39**, 1569–1573 (1980).
- Silver, P. J. & DiSalvo, J. *J. biol. Chem.* **254**, 9951–9954 (1979).
- Nonomura, Y. & Ebashi, S. *Biomed. Res.* **1**, 1–14 (1980).
- Kerrick, W. G. L. & Krasner, B. J. *appl. Physiol.* **39**, 1052–1055 (1975).
- Hellam, D. C. & Podolsky, R. J. *J. Physiol., Lond.* **200**, 807–819 (1969).
- Donaldson, S. K. B. & Kerrick, W. G. L. *J. gen. Physiol.* **66**, 427–444 (1975).

Progesterone inhibits membrane-bound adenylate cyclase in *Xenopus laevis* oocytes

Joëlle Finidori-Lepicard*, Sabine Schorderet-Slatkine†, Jacques Hanoune‡ & Etienne-Emile Baulieu*§

* U 33 Inserm and Faculté de Médecine, 78 rue du Général Leclerc, 94270 Bicêtre, France

† Département de Gynécologie et d'Obstétrique, Hôpital Cantonal, Genève, Switzerland

‡ U 99 Inserm, 51 avenue du Maréchal de Lattre de Tassigny, 94000 Créteil, France

Recent experimental evidence indicates that progesterone acts at the cell surface to trigger protein synthesis and to reinitiate the first meiotic division in *Xenopus laevis* oocytes^{1,2}. The steroid hormone is physiologically released by follicle cells surrounding oocytes in the ovaries, and this naturally occurring event can be reproduced *in vitro* by adding progesterone to the incubation medium. Recently, cyclic AMP has been implicated in the mechanism of progesterone action in oocytes^{3,4}; there was an almost immediate decrease in cyclic AMP concentration in oocytes after addition of progesterone *in vitro*, whether or not the oocytes were pretreated with cholera toxin^{5,6}. Adenylate cyclase in *X. laevis* oocytes is compartmentalized, >50% soluble and ~30% is found in the plasma membrane-containing fraction⁷. We report here that physiological concentrations of progesterone selectively inhibit membrane-bound adenylate cyclase activity, after addition to intact oocytes or in cell-free experiments; this specificity confirms the proposed membrane site of action for the hormone when reinitiating meiosis and is the first example of a 'direct' enzymatic action of a steroid (not by protein synthesis) related to a physiological function.

Adenylate cyclase activity in oocytes was measured in four subcellular fractions isolated by differential centrifugation (Fig. 1a), and was largely found in two compartments, a 10,000g pellet (P-10,000; 20–30%) and the cytosol (50–65%), a distribution similar to that observed in the early stages of male germ-cell development⁸. Both particulate and soluble forms of the enzyme appeared equally active in the presence of either

Mg²⁺ or Mn²⁺, but only the particulate activity was stimulated (two- to threefold) by 10 mM sodium fluoride. Electron microscopic examination of P-10,000 showed a uniform vesiculated membrane structure and enzyme patterns suggested an enrichment in plasma membrane. Incubation of oocytes with progesterone produced a rapid inhibition of the P-10,000-associated cyclase. Its specific activity was reduced by 65% while the soluble activity was unchanged. Cycloheximide had no effect on particulate cyclase inhibition. The facts that only the membrane-bound cyclase was inhibited by progesterone, and that more than half of the oocyte enzymatic activity is soluble, explain the limited decrease in total cyclic AMP observed in intact oocytes. Therefore, cholera toxin was used to magnify the progesterone-induced decrease of cyclic AMP in oocytes; after cholera toxin has been added to increase cyclic AMP levels (to 150–300% of control), the addition of progesterone produces an immediate decrease to basal level⁶. Adenylate cyclase determinations confirmed the dual nature of the enzyme in oocytes, as only the P-10,000 cyclase was stimulated by cholera toxin (5–10-fold), while this stimulation was reversed by addition of progesterone to the oocyte incubation medium just before cell fractionation (Fig. 1b).

To demonstrate that progesterone interacts directly with the plasma membrane cyclase activity, without an intermediary

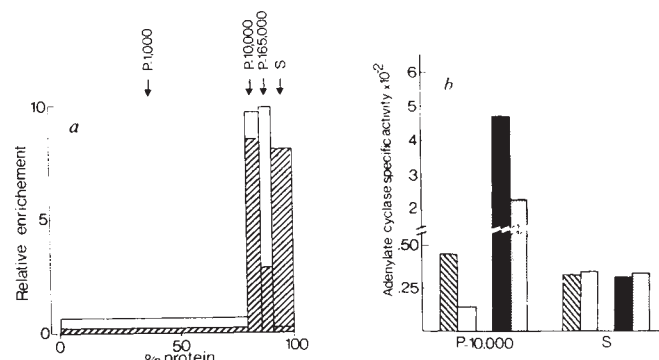


Fig. 1 Adenylate cyclase activity in subcellular fractions of untreated and progesterone- and/or cholera toxin-treated *X. laevis* oocytes. Fully grown oocytes, obtained as described elsewhere⁴, were washed and homogenized in 10 vol of 1 mM NaHCO₃, 3 mM EDTA, at 4°C, with a Dounce apparatus (pestle A, 10 strokes). The homogenate was centrifuged at 1,000g for 15 min; the pellet (P-1,000) contained most of the vitellus and melanosomes. The supernatant was centrifuged at 10,000g for 20 min and the P-10,000 pellet separated. The supernatant was again centrifuged at 165,000g for 15 min, and both the supernatant (S) and the pellet (P-165,000) were retained. All pellets were resuspended in the same buffer and enzymatic assays were performed immediately. Adenylate cyclase assays were done in 60 µl containing 1 mM [α -³²P]ATP (10⁶ c.p.m.), 3 mM MgCl₂, 1 mM EDTA, 1 mM cyclic AMP, 50 mM Tris-HCl pH 7.6, plus an ATP-regenerating system consisting of 25 mM phosphocreatine and 1 mg ml⁻¹ creatinine phosphokinase. The reaction was terminated by a modification¹⁴ of the procedure of White¹⁵. In these conditions, the degradation of ATP was negligible and the enzymatic activity was linear with time up to 60 min and linear with protein up to 60 µg per assay; this was also the case for inhibition experiments. Labelled cyclic AMP was isolated according to Rachamadrán¹⁶. The yield was calculated by previous addition of cyclic [8-³H]AMP; it was unchanged when the experiment was performed in the presence of steroids, eliminating a significant stimulation of the phosphodiesterase(s) by the steroids. 5'-nucleotidase activity was estimated as described elsewhere¹⁷. The distribution pattern of adenylate cyclase (■) in oocytes is shown in a, superimposed on that of 5'-nucleotidase (□). Enzyme activities and protein concentrations were measured in the whole homogenate and in the separated fractions. Relative enrichment for adenylate cyclase is the ratio of specific activity in each fraction divided by specific activity of the whole homogenate. This parameter is plotted against the percentage of the whole homogenate protein recovered in each fraction. The area of the bars is proportional to the recovery of the enzyme in the corresponding fractions. The ordinate measures the degree of purification achieved over the whole homogenate value. The same representation is used for 5'-nucleotidase. b, Adenylate cyclase specific activities in P-10,000 and S fractions of untreated oocytes (cross-hatched bars), oocytes preincubated with 30 µM progesterone for 90 s (open bars), oocytes preincubated with 50 pmol cholera toxin for 6 h (solid bars), and oocytes incubated with 30 µM progesterone for 90 s at the end of the preincubation with cholera toxin (stippled bars). Cyclase activity is expressed as pmol of cyclic AMP formed in 15 min per mg protein at 32°C. Protein and 5'-nucleotidase values were identical in control and progesterone/cholera toxin-treated oocytes (data not shown).

§ To whom reprint requests should be addressed.

state involving a metabolic change, we performed cell-free experiments. We found that P-10,000 adenylate cyclase from control oocytes was inhibited by addition of progesterone in the enzymatic assay at 20 °C—the physiological temperature for the maturation process—as well as at 32 °C, the optimal temperature for the cyclase assay. There was no lag phase and the decrease in cyclase activity persisted for at least 60 min (in agreement with the prolonged reduction in cyclic AMP level observed in intact oocytes⁶). Progesterone-induced inhibition of cyclase activity was dose dependent: $K_i \sim 0.1 \mu\text{M}$ (that is about one order of magnitude lower than the half-maximal effective dose of progesterone when initiating meiosis, as if some trapping or inactivation of the hormone in intact oocytes may be eliminated in the cell-free system). In most cases, maximum inhibition (at 10 μM) reached 80% of control activity. Progesterone reduced the V_{max} of the enzyme without changing the apparent affinities for MgATP and MnATP substrates, and for the activators free Mg^{2+} and free Mn^{2+} (data not shown). Finally, cyclase activity was entirely recovered after extensive membrane washing, which demonstrates the reversibility of progesterone inhibition. Adenylate cyclase activities stimulated by NaF or guanylyl-5'-yl imidodiphosphate (Gpp(NH)p, a GTP analogue not hydrolysed to GDP by GTPases in membranes) were also inhibited by progesterone, which also reversed enzyme stimulation in P-10,000 obtained from oocytes preincubated with cholera toxin (Fig. 2). In all these assay conditions, inhibition by progesterone was dose dependent with an apparent K_i similar to that found in basal conditions. This effect of progesterone seemed to be different from the GTP-dependent inhibitions so far described, which are prevented by stable GTP analogues and generally abolished by cholera toxin pretreatment⁹. Progesterone inhibition did not seem to be mediated by an adenosine effect as it was not suppressed by adenosine desaminase in the assay; adenosine and progesterone inhibitions of cyclase in oocyte membranes were additive⁷.

Furthermore, progesterone-induced inhibition of adenylate cyclase does not seem dependent on Ca^{2+} concentrations. In intact oocytes, progesterone effect was observed in Barth's medium containing 1 mM CaCl_2 , while in cell-free experiments steroid inhibition occurred in the presence of 1 mM EDTA. *In vitro*, Ca^{2+} -induced inhibition of cyclase ($K_i = 0.5 \text{ mM}$) and progesterone inhibition were additive, which suggests that these

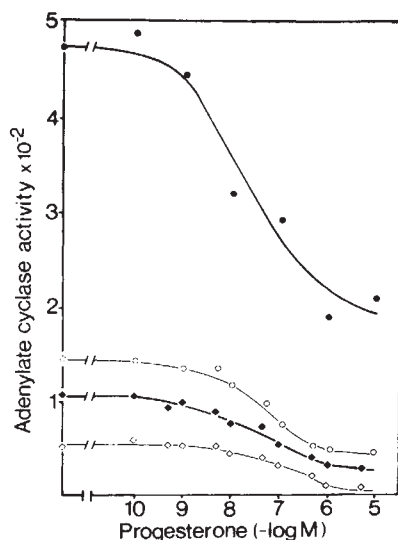


Fig. 2 Inhibition by progesterone of membrane adenylate cyclase activity in *X. laevis* oocytes treated with cholera toxin or untreated. The P-10,000 fraction was obtained from untreated *X. laevis* oocytes ($\diamond$, $\bullet$, $\circ$) or from oocytes preincubated for 6 h with 50 pM cholera toxin ($\bullet$). Progesterone was added in the enzymatic assay, with no other addition ($\diamond$, $\bullet$), or with 10 mM sodium fluoride ($\bullet$) or 1 μM Gpp(NH)p ($\circ$). The reaction was initiated by addition of P-10,000 (50 μg protein per assay) to the enzymatic assay medium. Adenylate cyclase activity is expressed as pmol of cyclic AMP formed in 15 min per mg protein.

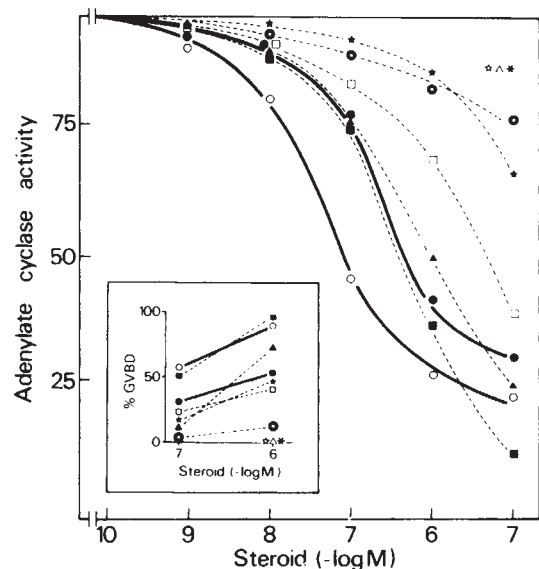


Fig. 3 Membrane adenylate cyclase of *X. laevis* oocytes: specificity of steroid inhibition. The P-10,000 fraction was incubated with 1 μM Gpp(NH)p and steroids. Adenylate cyclase activity at each concentration of steroid tested is expressed as per cent of the activity measured in the absence of hormone. Percentage of germinal vesicle breakdown (GVBD) is indicated in the inset. It was routinely determined on 30–50 oocytes after 8 h incubation by scoring the appearance of the 'white spot' at the animal pole. Experiments were done exactly as described elsewhere⁴. The steroids tested were: progesterone ($\circ$); progesterone + 10 μM 17 α -ethynyl-oestradiol ($\bullet$); testosterone ($\blacksquare$); cortisone ($\blacktriangle$); dehydroepiandrosterone ($\square$); hydrocortisone ($\star$); 19-nortestosterone ($\blacklozenge$); oestrone ($\star$); 17 α -ethynyl-oestradiol ($\triangle$) and progesterone ($\ast$).

two inhibitors use different mechanisms. However, we cannot exclude the possibility that changes in Ca^{2+} contribute to the inhibition of adenylate cyclase activity in intact oocytes.

The inhibitory steroid effect showed the same hormonal specificity as that reported for maturation in intact oocytes. In the P-10,000 cell-free system, testosterone and cortisone, which are almost as efficient in promoting meiosis as progesterone, were also the most active inhibitors of adenylate cyclase ($K_i = 0.1 \mu\text{M}$). Dehydroepiandrosterone, cortisol and 19-nortestosterone only reached 2/3–1/3 of the progesterone inhibitory activity, with an apparent $K_i \sim 1 \mu\text{M}$; oestrone, 17 α -ethynyl-oestradiol and progesterone (1,3,5-pregnen-3-ol, 20-one) were not active. Moreover, we found that 17 α -ethynyl-oestradiol, which antagonizes progesterone-induced meiosis apparently in a competitive manner¹⁰, displaced to the right by one order of magnitude the dose-dependent curve of cyclase inhibition by progesterone (Fig. 3).

These results provide the first example of a direct enzymatic action of steroids in a physiological mechanism, the enzyme being a typical plasma membrane component. Intriguing studies with glutamate dehydrogenase demonstrated that steroids can have allosteric effects on the enzyme¹¹, and it has been suggested that this could be an important mechanism for hormone action¹². However, the lack of hormonal specificity made this model system irrelevant to a physiological system. The remarkable correlation between steroid action on meiosis and on membrane adenylate cyclase inhibition strongly suggests that we are dealing with a biochemical event relevant to the biological phenomenon. Experiments are now being done to determine whether other steroid systems show characteristics similar to those observed in oocytes¹³.

Received 9 April; accepted 5 May 1981.

- Smith, L. D. & Ecker, R. E. *Devl Biol.* **25**, 233–247 (1971).
- Godeau, J. F., Schorderet-Slatkine, S., Hubert, P. & Baulieu, E. E. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2353–2357 (1978).
- Maller, J. L. & Krebs, E. G. *J. biol. Chem.* **252**, 1712–1718 (1977).

4. Schorderet-Slatkine, S., Schorderet, M., Boquet, P., Godeau, F. & Baulieu, E. E. *Cell* **15**, 1269–1275 (1978).
5. Maller, J. L., Butcher, F. R. & Krebs, E. G. *J. biol. Chem.* **254**, 579–582 (1979).
6. Schorderet-Slatkine, S., Baulieu, E. E., El Etr, M. & Schorderet, M. in *International Cell Biology 1980–1981* (Schweiger, H. G. ed.) 860–871 (Springer, Berlin, 1981).
7. Finidori-Lepicard, J., Schorderet-Slatkine, S., Hanoune, J. & Baulieu, E. E. (in preparation).
8. Braun, T. & Dods, R. F. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1097–1101 (1975).
9. Jakobs, K. H., Aktories, K., Nasch, P., Saur, W. & Schultz, G. in *Hormones and Cell Regulation* Vol. 4 (eds Dumont, J. & Nunez, J.) 89–106 (Elsevier/North-Holland, Amsterdam, 1980).
10. Baulieu, E. E., Godeau, F., Schorderet, M. & Schorderet-Slatkine, S. *Nature* **275**, 593–598 (1978).
11. Tomkins, G. M. & Maxwell, E. S. A. *Rev. Biochem.* **32**, 677–708 (1963).
12. Monod, J. *Endocrinology* **78**, 412–425 (1966).
13. Baulieu, E. E. in *Central Regulation of the Endocrine System* (eds Fuxe, K., Hökfelt, T. & Luft, R.) 239–260 (Plenum, New York, 1978).
14. Hanoune, J., Stengel, D., Lacombe, M. C., Feldmann, G. & Coudrier, E. *J. biol. Chem.* **252**, 2039–2045 (1977).
15. White, A. A. *Meth. Enzym.* **38**, 41–46 (1974).
16. Ramachandran, J. *Analyt. Biochem.* **43**, 227–239 (1971).
17. Pecker, F., Duvaldestin, P., Berthelot, P. & Hanoune, J. *Clin. Sci.* **57**, 313–325 (1979).

Differences between oestrogen receptor activation by oestrogen and antioestrogen

Henri Rochefort & Jean-Louis Borgna

Unité d'Endocrinologie Cellulaire et Moléculaire, U 148 INSERM, 60, rue de Navacelles, 34100 Montpellier, France

Triphenylethylene antioestrogens such as tamoxifen, nafoxidine and Ci 628 specifically inhibit oestrogen action at the target cell level¹, probably by interacting with the oestrogen receptor (ER) and competitively displacing oestrogens from their binding sites. It is not clear, however, why these ligands are less biologically active than oestrogens when they bind to the ER, as no reliable difference has been found either in the binding affinity of these two series of ligands to the ER or in their ability to translocate the ER to the nucleus^{1,2}. In fact, these antioestrogens are transformed *in vivo* into hydroxylated metabolites^{3–5} which display a better antioestrogenic activity than the injected compound and at least the same high affinity as oestradiol for the ER⁶. With the aim of finding an *in vitro* criterion to predict the agonistic or antagonistic properties of ER ligands, we have stabilized the ER in its 'native' or non-activated form by the use of molybdate^{7,8} and have compared the binding of oestradiol (E₂) and of 4-hydroxytamoxifen (OHT), an active metabolite of tamoxifen, to the molybdate-treated and to the activated ER. We report here that molybdate prevented the DNA binding and the 4S to 5S transformation of the ER bound to both ligands, and that it increased the dissociation rate of oestrogens but not that of antioestrogens. Moreover, in the absence of molybdate, receptor activation by heating decreased the dissociation rate of E₂ but not that of OHT. We conclude that a difference exists between the ER activation triggered by oestrogens and antioestrogens and propose that antioestrogens are acting as allosteric ligands of the ER.

We have looked at three series of *in vitro* criteria for receptor activation. The first classical criterion is the increased affinity of the receptor for DNA and nuclei. The ability of lamb uterine ER to interact *in vitro* with calf thymus DNA was evaluated by ultracentrifugation in sucrose⁶ and metrizamide⁹ gradients. In the absence of molybdate, both the ER–E₂ and the ER–OHT complexes bound to DNA, confirming that OHT also activate the ER for binding to DNA *in vitro*⁶ and for translocating the ER into the nucleus¹. With molybdate, the DNA binding of the ER was totally inhibited whether ER was bound to an antioestrogen (OHT) or to an oestrogen (E₂). A second *in vitro* test for ER activation is the E₂-induced 4S to 5S transformation¹⁰ which is also prevented by molybdate. However, both OHT and E₂ were able to transform the cytosol 4S ER into a 5S ER in the immature rat uterus and both processes were prevented by molybdate.

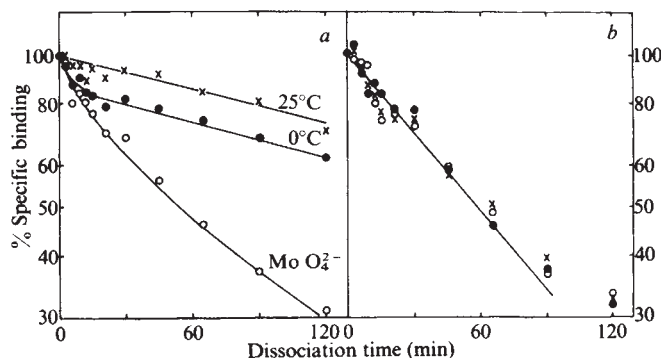


Fig. 1 Dissociation rate of oestradiol (a) and 4-hydroxytamoxifen (b) from the oestrogen receptor before and after heat activation. Lamb uterine cytosol (2 mg protein per ml) prepared in TE buffer (Tris-HCl 10 mM, EDTA 1.5 mM, pH 7.4) in the presence (○) or absence (●, ×) of 10 mM MoO₄²⁻ was incubated for 2 h with 5 nM ³H-oestradiol (E₂) (CEA: specific activity 55 Ci mmol⁻¹) or ³H-4-hydroxy-*trans*-tamoxifen (OHT) (specific activity 15.8 Ci mmol⁻¹) prepared biologically from ³H-tamoxifen or provided by ICI. One part of the cytosol without MoO₄²⁻ was activated by a 30-min treatment at 25°C, the other part was maintained at 0°C, the dissociation rate was then determined at 25°C as described elsewhere⁶, following the addition of 1 μM unlabelled E₂ at time 0. At the indicated times, aliquots were removed and cooled at 0°C. The specific binding to the ER was determined after 90 min incubation with charcoal suspension⁶; the non-specific binding was evaluated in a parallel experiment performed with 1 μM unlabelled E₂. Results were normalized by taking the pre-dissociation value as 100%. The complexes were stable during the course of the assay. × (25°C), preactivated ER; ● (0°C), non-preactivated ER; ○ (MoO₄²⁻), molybdate-treated ER.

Table 1 Effect of molybdate on the dissociation rate (k^-) of various oestrogens and antioestrogens from the oestrogen receptor

a, Species specificity (k^- , $\times 10^{-4}$ s ⁻¹)						
	Calf	Lamb	Rat	Mouse	Chicken	MCF ₇
Temperature (°C)	20	20	20	20	15	20
E ₂						
Control	0.47	0.14	0.64	1.27	0.68	0.51
+Molybdate	1.10	0.28	2.40	4.01	1.75	1.04
% Increase	135	100	275	215	155	105
OHT						
Control	0.56	0.47	0.69	0.88	0.20	1.92
+Molybdate	0.56	0.47	0.69	0.88	0.20	1.92
% Increase	0 ± 10	~0	~0	~0	~0	~0
b, Ligand specificity (k^- , $\times 10^{-4}$ s ⁻¹)						
	Temperature (°C)			Control	+Molybdate	% Increase
E ₂	20			0.14	0.28	100
E ₁	20			6.60	11.91	80
E ₃	20			0.31	0.68	120
5-Adiol	0			0.63	1.13	80
OHT	20			0.47	0.47	≈0
T	0			2.30	2.30	≈0
Ci 628	0			0.85	0.85	≈0

a, Species specificity: the dissociation rate constants of E₂ and OHT from the cytosol ER of calf, lamb, rat and mouse uterus, of chicken oviduct and of the MCF₇ human breast cancer cell line were determined at 20°C or 15°C (chicken), following heat 'activation' in the presence or absence of 10 mM sodium molybdate as described in Fig. 2 legend. The k^- values calculated after correction for nonspecific binding and their percentage increase due to molybdate are presented. For OHT studies, the difference between the two linear regressions drawn through the points obtained with and without molybdate was zero or less than 10%. We have therefore represented only the mean values. b, Ligand specificity: the k^- from the cytosol ER of lamb uterus has been determined at 20 or 0°C as described in Fig. 2 legend for various oestrogens (oestradiol = E₂; oestrone = E₁; oestriol = E₃; androsta-5-ene-3 β ,17 β -diol = 5-Adiol) and antioestrogens (Ci 628; tamoxifen = T; 4-hydroxytamoxifen = OHT).

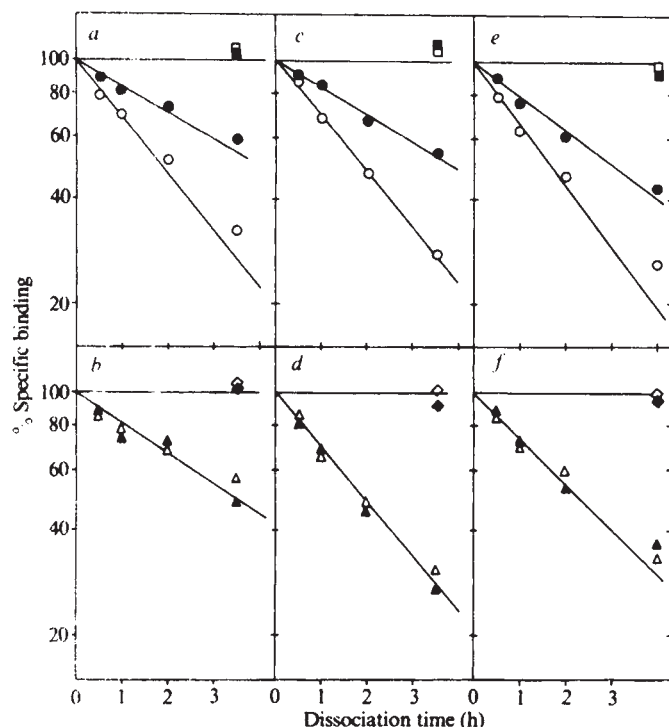


Fig. 2 Effect of molybdate on the dissociation rate of oestrogens and antioestrogens from the oestrogen receptor. *a, b*, Calf uterine cytosol prepared with (○, △) or without (●, ▲) 10 mM MoO_4^{2-} was incubated for 2 h with 5 nM $^3\text{H-E}_2$ (○, ●) or $^3\text{H-OHT}$ (△, ▲). The cytosol ER was then heated for 30 min at 25 °C and the dissociation rate at 20 °C was evaluated as described previously⁶. Each value was corrected for nonspecific binding. The stability of the complexes was assayed separately (□, ■, ◇, ◆). Results were normalized by taking the pre-dissociation value as 100%. *c, d*, Same experiment with cytosol of the MCF₇ human breast cancer cells. The ER was activated at 20 °C. *e, f*, Dissociation rate at 0 °C from lamb uterine cytosol incubated with 20 mM $^3\text{H-androsta-5-ene, 3}\beta,17\beta\text{-diol}$ (○, ●) (specific activity 58.6 Ci mmol⁻¹) or $^3\text{H-Ci 628}$ (△, ▲) (22.5 Ci mmol⁻¹), both from the Radiochemical Centre, in the presence or absence of 2 μM diethylstilboestrol to assay nonspecific binding. After heating for 30 min at 25 °C, the dissociation of the ligands at 0 °C was assayed by adsorption with charcoal for 30 min at 0 °C.

Recently, a third criterion has been proposed^{8,11,12}, the slower dissociation rate of oestrogens from the activated ER than from the 'native' non-activated ER. We therefore compared the dissociation rate at 25 °C for $^3\text{H-E}_2$ and $^3\text{H-OHT}$ (Fig. 1). As already described², the dissociation rate from the cytosol ER varied according to whether the cytosol was treated at 0 °C or was preheated at 25 °C to activate the complex. In the initially non-activated ER, the dissociation of $^3\text{H-E}_2$ was biphasic. The first slope was more rapid than the second which corresponded to the activated complex formed during dissociation. When the cytosol ER had been preactivated, the dissociation was monophasic and parallel to the second slope. Conversely, with $^3\text{H-OHT}$, no difference of dissociation could be seen whether or not the cytosol was preheated at 25 °C. This suggested a difference of activation between the ER- E_2 and ER-OHT complexes. Because the measurement of the dissociation rate for these high-affinity ligands required elevated temperature which also activated the receptor, we then tried to stabilize the native non-activated receptor by using molybdate.

As seen in Fig. 1, the ER- E_2 complex continued to dissociate at the initial rate in the presence of molybdate, which apparently prevents the activation process. Conversely, we found no difference for the dissociation of $^3\text{H-OHT}$ from the ER. This seems to be a general phenomenon because, in all species tested, the dissociation of E_2 was more rapid from the molybdate-stabilized ER, whereas that of OHT was generally not modified by molybdate (Fig. 2, Table 1). In the rat and mouse uterus where antioestrogens are partial agonists¹³, the dissociation rate of OHT was also unchanged by molybdate. However, in these species, the metabolite(s) responsible for the agonist activity are unknown. Other oestrogens—oestrone, oestriol and androsta-5-ene, $3\beta,17\beta\text{-diol}$ (5-Adiol)—which bind to the ER with a K_D of 6 nM (ref. 14), also dissociated more rapidly from the ER stabilized with molybdate than from the untreated ER (Table 1b). Conversely, the dissociation rate of the other antioestrogens, tamoxifen and Ci 628, was not altered by molybdate in lamb uterus.

These kinetic experiments indicate an increased affinity of E_2 for the ER during its activation, in contrast to a relatively constant affinity of antioestrogens for the ER. Thus, when antioestrogens compete with E_2 for binding to the ER, the competition should be more effective initially and should diminish progressively with time, E_2 displacing the antioes-

trogen from the ER. We have previously observed such a displacement¹⁵. To specify whether this displacement was due to the activation of the ER or simply to a lack of equilibrium between the two ligands and the ER¹⁶, we have tested the effect of molybdate during competition of $^3\text{H-E}_2$ binding by tamoxifen. Figure 3 shows that the competitive efficiency of tamoxifen was increased by the addition of molybdate. However, the time-dependent increase of E_2 binding was still observed, suggesting that it was due partly to a lack of equilibrium and partly to receptor activation.

The action of non-steroidal antioestrogens is mediated either by the ER as classically proposed or by other receptor entities, such as the recently described¹⁷ protein binding specifically anti-oestrogens. In support of the latter hypothesis, we have found in MCF₇ breast cancer cells, which do not metabolize tamoxifen or OHT, that OHT is more active than tamoxifen in inhibiting the induction by E_2 of a 46,000 molecular weight protein¹⁸ and in preventing cell growth¹⁹. However, the finding that the effect of antioestrogens in MCF₇ cells²⁰ and in chick liver²¹ is fully and rapidly reversed by the addition of oestrogens, strongly suggests that these antioestrogens are simply acting by competing with oestrogens on the ER. Nevertheless, it is not clear why the ER-antioestrogen complex is not fully active or is in some cases totally unable to induce oestrogen-specific proteins. The simplest explanation is that the E_2 -induced activation of the ER is different from that triggered by an antioestrogen. This altered activation of the ER by antioestrogen might therefore be detectable before the nuclear translocation step, even though its consequence will be located at sites of interaction with chromatin. An alternative hypothesis is an altered interaction with chromatin of a normally activated ER-antioestrogen

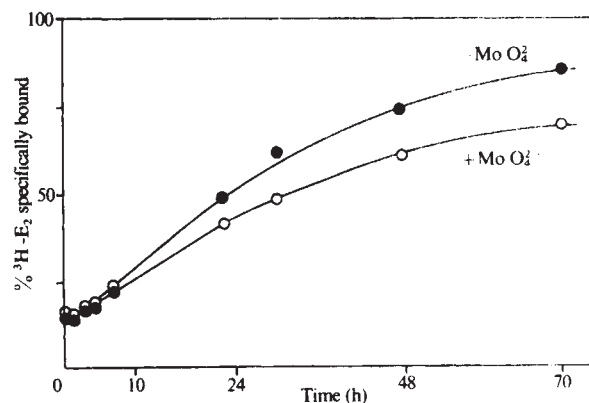


Fig. 3 Time-dependent variation of competitive efficiency of tamoxifen and effect of molybdate. Lamb uterine cytosol was incubated in the presence (○) or absence (●) of molybdate with 2 nM $^3\text{H-E}_2 \pm 0.5 \mu\text{M}$ tamoxifen for the indicated time periods. The $^3\text{H-E}_2$ -specific binding was determined by charcoal assay⁶ and the binding obtained in the presence of tamoxifen was expressed as a percentage of the non-inhibited binding at each point.

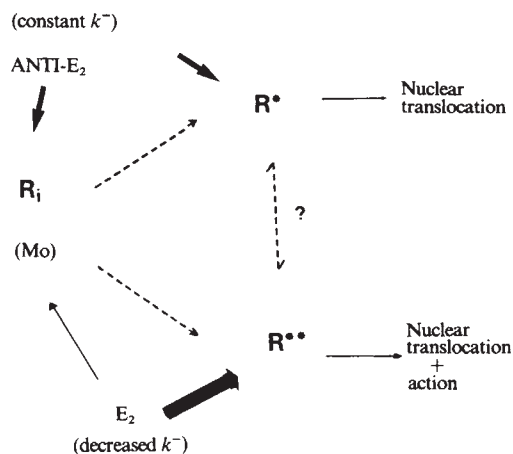


Fig. 4 A model for two types of oestrogen receptor activation induced by oestrogens or antioestrogens. The native ER (R_i) is in the cytosol. After interaction with antioestrogens (anti- E_2), it is activated for nuclear translocation (R^*) but not for a complete action. After interaction with oestrogens (E_2), the ER is fully activated for nuclear translocation and for action (R^{**}). We propose to differentiate between the two types of activation by measuring the *in vitro* dissociation rate of ligands from the molybdate-stabilized R_i and the activated R . Whereas oestrogens dissociate at a slower rate from R^{**} than from R_i , antioestrogens dissociate at the same rate from both R^* and R_i . The width of the full arrows indicates the relative affinity of the ligands for each receptor state.

complex. This could be envisaged, for instance, as a direct interaction of the free drug with DNA or protein in chromatin. Our results strongly support the first possibility. This difference of activation by antioestrogens is detected before the translocation step by simply analysing the dissociation rate of the antioestrogen from the ER with and without molybdate. The consequence of this impeded activation is probably located at the chromatin level, where the ER is thought to trigger the first biochemical responses.

As the oestrogens but not the antioestrogens are able to transform the ER–ligand complex into a more slowly dissociating form, although both ligands can induce the ER nuclear translocation, we propose two kinds of ER activation (Fig. 4). The first (R^*) would be obtained after binding of any ligand, oestrogen or antioestrogen, and would trigger a covalent or conformational change of ER allowing its nuclear localization. The second type of activation (R^{**}) would be obtained only after binding to oestrogens and would induce a more complete modification of the ER, giving a full metabolic response. A practical *in vitro* criterion for a full biological activation of the receptor would therefore be a modification of the ER binding site such that the dissociation rate of agonists, but not antagonists, is reduced. It is thus possible to discriminate between oestrogen agonists and antagonists by simply comparing the dissociation rate constants of the ligands from the molybdate-stabilized (non-activated) and from the activated forms of the ER. Note that it is not the actual value of the k^- , but rather its variation during ER activation, which characterizes an active oestrogen. In this model, the affinity of the antioestrogen for the activated and non-activated ER would be similar, whereas the affinity of oestrogen would be higher for the activated form of the ER. We propose that antioestrogens act as allosteric²² ligands which stabilize the ER in inactive (R_i) or partially activated form (R^*), whereas oestrogens stabilize the receptor in a fully activated form (R^{**}). These results further support the generally accepted idea that antioestrogens act via the ER and not via a distinct receptor molecule specific for antioestrogens.

We thank Dr Todd (ICI) for providing 3H -tamoxifen and 3H -4-hydroxytamoxifen, Dr D. Bullock for critical reviewing of the manuscript, S. Ladrech for technical assistance and E. Barrié

for other help. This work was supported by grants from CNRS (ATP 41–61), INSERM (CRL 79.5.235.4) and the Association pour le Développement de la Recherche sur le Cancer.

Received 10 February; accepted 12 May 1981.

- Clark, J. H. *et al.* (eds) *Dahlem Conf., Life Sci. Res. Rep.* **3**, 147–169 (1976).
- Rocheftort, H., Garcia, M. & Borgna, J. L. *Biochem. biophys. Res. Commun.* **88**, 351–357 (1979).
- Katzenellenbogen, B. S., Katzenellenbogen, J. A., Ferguson, E. R. & Krauthammer, N. J. *biol. Chem.* **253**, 697–707 (1978).
- Borgna, J. L. & Rocheftort, H. *J. biol. Chem.* **256**, 859–868 (1981).
- Jordan, V. C., Collins, M. M., Rowsby, L. & Prestwich, G. J. *Endocrinology* **75**, 305–357 (1979).
- Borgna, J. L. & Rocheftort, H. *Molec. cell. Endocr.* **20**, 71–86 (1980).
- Mauk, L. A. & Notides, A. C. *62nd A. Meet. Endocr. Soc.*, Abstr. 209 (1980).
- Shyamala, G. & Leonard, L. J. *biol. Chem.* **255**, 6028–6031 (1980).
- Baskevitch, P. P. & Rocheftort, H. *Molec. cell. Endocr.* **22**, 195–210 (1981).
- Notides, A. C. & Nielsen, S. J. *J. biol. Chem.* **249**, 1866–1873 (1974).
- Raynaud, J. P., Bouton, M. M. & Ojasoo, T. *Trends pharmac. Sci.* **3**, 324–327 (1980).
- Weichman, B. M. & Notides, A. C. *Endocrinology* **106**, 434–439 (1980).
- Black, L. J. & Goode, R. L. *Life Sci.* **26**, 1453–1458 (1980).
- Garcia, M. & Rocheftort, H. *Endocrinology* **104**, 1797–1804 (1979).
- Rocheftort, H. & Capony, F. *Biochem. biophys. Res. Commun.* **75**, 277–285 (1977).
- Aranyi, P. *Biochim. biophys. Acta* **628**, 220–227 (1980).
- Sutherland, R. L., Murphy, L. C., San Foo, M., Green, M. D. & Whybourne, A. M. *Nature* **288**, 273 (1980).
- Westley, B. R. & Rocheftort, H. *Cell* **20**, 353–362 (1980).
- Rocheftort, H., Borgna, J. L., Coezy, E., Vignon, F. & Westley, B. in *Non-Steroidal Antiestrogens* (eds Sutherland, R. L. & Jordan, V. C.) (Academic, Sydney, 1980).
- Lippman, M., Bolan, G. & Huff, K. *Cancer Res.* **36**, 4595–4601 (1976).
- Capony, F. & Williams, D. L. *Endocrinology* **19**, 2219–2226 (1980).
- Bullock, L. P., Bardin, C. Q. & Sherman, M. *Endocrinology* **103**, 1768–1782 (1978).

Human transforming growth factors induce tyrosine phosphorylation of EGF receptors

Fred H. Reynolds Jr*, George J. Todaro†, Charlotte Fryling† & John R. Stephenson†

*Carcinogenesis Intramural Research Program, Frederick Cancer Research Center, Frederick, Maryland 21701, USA

†Laboratory of Viral Carcinogenesis, National Cancer Institute, Frederick, Maryland 21701, USA

Cultured cell lines of human tumour origin as well as cells transformed by various RNA tumour viruses secrete low molecular weight polypeptide transforming growth factors (TGFs)^{1,2}. In addition to competing with epidermal growth factor (EGF) for binding to its cellular receptor, TGFs can transform morphologically fibroblast and epithelial cells in culture^{1–3}. In view of accumulating evidence that tyrosine phosphorylation activity is associated with the transforming genes of various tumour viruses^{4–13}, we determined whether phosphotyrosine levels were elevated in these human tumour cells. We show here that TGFs produced by human tumour cells induce phosphorylation of specific tyrosine acceptor sites in the 160,000-molecular weight (160 K) EGF receptor.

The demonstration of increased levels of phosphotyrosine-containing proteins in cells transformed by avian^{5–8} and mammalian^{9–13} RNA tumour viruses raises the question of whether regulatory pathways which involve tyrosine phosphorylation may also have significance in the aetiology of spontaneous and chemically induced tumours. We therefore examined phosphotyrosine levels in several established human tumour cell lines. Of six independently derived lines analysed, none showed overall phosphotyrosine concentrations that were significantly greater than those of untransformed control Fisher rat embryo cells (Table 1). Similarly, we (data not shown) and Sefton *et al.*⁵ have analysed many non-virally transformed mammalian cells of other than human origin and found that the phosphotyrosine-containing protein levels are comparable with those of untransformed cells. In analogous assay conditions, as reported elsewhere^{5,9,12,13}, cells transformed by either Gardner feline sarcoma virus (FeSV) or Abelson murine leukaemia virus (AbLV) show 8–10-fold increases in phosphotyrosine levels,

Table 1 Phosphotyrosine levels in human tumour and Fisher rat cells after exposure to either EGF or human TGF

Cell line	Phosphotyrosine (% of total phosphoamino acids)		
	Untreated	EGF-stimulated	TGF-stimulated
Human tumour:			
A431 carcinoma	<0.2	2.6	2.4
A673 rhabdomyosarcoma	<0.2	<0.2	<0.2
A2058 melanoma	0.4	0.4	<0.2
A875 melanoma	<0.2	0.3	NT
JARC2 carcinoma	<0.2	<0.2	NT
A204 rhabdomyosarcoma	<0.2	<0.2	NT
Fisher rat:			
Control	<0.2	<0.2	<0.2
AbLV-transformed	1.9	2.0	NT
G-FeSV-transformed	1.7	1.9	1.8
M-MSV-transformed	<0.2	<0.2	NT

Cell lines were cultured in Dulbecco's modification of Eagle's medium as described previously¹. Phosphotyrosine determinations were performed by incubation of cells in medium containing ^{32}P -orthophosphate (1.0 mCi ml^{-1}) and as indicated above, either EGF ($0.1\text{ }\mu\text{g ml}^{-1}$) or transforming growth factor (TGF $10\text{ }\mu\text{g ml}^{-1}$), partially purified from medium of A673 cultures. ^{32}P -labelled proteins were prepared, hydrolysed and phosphoamino acids subjected to two-dimensional separation on cellulose TLC plates as previously described¹³. Individual ^{32}P -labelled phosphoamino acids, including phosphoserine, phosphothreonine and phosphotyrosine, were scraped off TLC plates and radioactivity quantified by liquid scintillation counting. Results are expressed as ^{32}P label in phosphotyrosine as a percentage of label in total phosphoamino acids. NT, not tested.

while the levels of phosphotyrosine in Moloney sarcoma virus (MSV)-transformed cells are similar to those in control cell lines.

In view of the phenotypic response of embryo fibroblasts to TGFs and EGF-stimulated tyrosine-specific protein kinase activity associated with the EGF membrane receptors¹⁴, we examined phosphotyrosine levels in human TGF- and mouse EGF-treated cells. As shown in Table 1, the human tumour line with the greatest concentration of available EGF receptors, A431 (refs 15, 16), showed a pronounced increase in total phosphotyrosine in response to either EGF or TGF, isolated from the culture fluids of the A673 rhabdomyosarcoma cell line¹. The overall extents of tyrosine phosphorylation in these growth factor-treated cells were comparable with those found in RNA tumour virus-transformed cells. Concentrations of phosphotyrosine in other cell lines, including A673 and A2058, which themselves are high-level producers of TGFs¹, remained unaltered after exposure to either human TGF or mouse EGF. Finally, the already elevated levels of phosphotyrosine in Gardner FeSV- and AbLV-transformed rat cells showed no further increase in response to either growth factor.

The A431 cell substrate(s) phosphorylated in response to TGFs were identified by an *in vitro* extract labelling procedure, which has been previously shown to result in preferential phosphorylation of tyrosine acceptor sites¹⁷. Figure 1 shows that phosphorylation of a 160,000- M_r major protein was observed after exposure of A431 cells to EGF for 1 min (lane a), but was not seen in control cells (lane c). ^{32}P -labelling of this substrate, designated P160, was observed at EGF concentrations as low as $0.01\text{ }\mu\text{g ml}^{-1}$ (Fig. 1, lane k), and was maximal at $1.0\text{ }\mu\text{g ml}^{-1}$ (Fig. 1, lane m). On the basis of both M_r and immunoprecipitation by an antiserum which specifically blocks binding of ^{125}I -labelled EGF to the previously described 160 K EGF membrane receptor, this phosphorylated cellular substrate was thought to be the EGF receptor itself. Exposure of A431 cells for 1 min to $100\text{ }\mu\text{g ml}^{-1}$ of either A673 human tumour-derived TGF (lane e) or sarcoma growth factor (SGF) isolated from supernatant fluids of MSV-transformed cells² (lane g) resulted in P160 phosphorylation comparable with that observed for higher levels of EGF ($1.0\text{ }\mu\text{g ml}^{-1}$). The higher specific activity of EGF as measured by stimulation of P160 phosphorylation or inhibition of ^{125}I -labelled EGF binding reflects the fact that it was fully purified whereas the TGF and SGF used here were only 1% pure. In experiments in which the length of exposure to EGF or TGF was increased to 10 min before extract pre-

paration, the extent of ^{32}P -labelling of P160 was considerably reduced and a second ^{32}P -labelled substrate of $\sim 35,000\text{ }M_r$ was observed. Analysis of Fisher rat cells for growth factor-induced phosphorylation of P160 showed no change in response to either EGF or TGF (Fig. 1, lanes f, h). Several lower molecular weight (20,000–40,000) proteins were phosphorylated in extracts of TGF- and SGF-treated rat cells but their significance is unknown. Other growth stimulatory factors including phorbol ester derivatives and insulin, when tested for stimulation of EGF receptor phosphorylation, had no effect in these assay conditions. Moreover, pretreatment of cells with phorbol ester derivatives in conditions previously shown to modulate EGF binding¹⁸ by converting available sites from high to low affinity^{19,20}, did not affect phosphorylation of P160 in response to either EGF or TGF.

Both the human and MSV-transformed mouse cell-derived factors have been further purified by HPLC²¹. Individual column fractions were assayed for inhibition of EGF binding, transformation of cultured rat embryo fibroblasts and stimulation of EGF receptor phosphorylation. All three activities co-purified in a region of a column representing <2.0% of the total input protein which argues strongly that EGF receptor-induced phosphorylation was due to the respective TGFs. In other studies, heating these factors in conditions (100°C , 5 min) which inactivate protein kinases neither destroyed the ability of TGFs to bind to EGF receptors nor altered P160-induced phosphorylation. Thus, the EGF receptor phosphorylation requires a kinase associated with the receptor. Binding of EGF, SGF, or TGF to this receptor alters its properties to give phosphorylation of the receptor protein.

These findings identify a substrate, P160, which is labelled in the presence of $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ in extracts of growth factor-treated but not untreated control cells. To test whether these extracts represented tyrosine acceptor sites, the 160 K EGF receptor, phosphorylated in response to EGF or to TGF, using *in vitro* extract labelling conditions, was eluted from SDS-polyacrylamide gel electrophoresis gels by trypsinization, and subjected to phosphoamino acid analysis. Figure 2 shows that

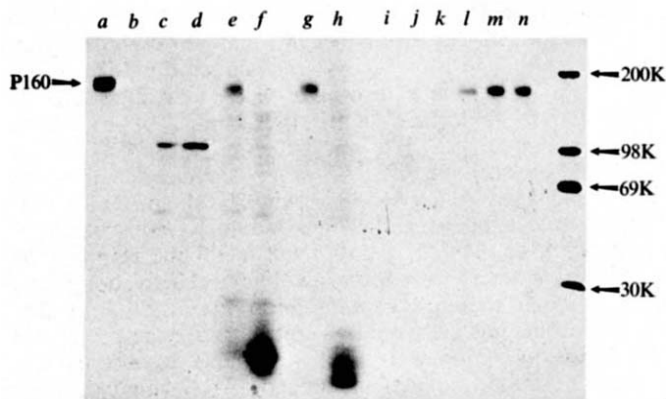


Fig. 1 SDS polyacrylamide gel electrophoresis (SDS-PAGE) analysis of substrates phosphorylated in A431 cell extracts in response to treatment with either mouse EGF or human TGF. A431 human tumour cells (a,c,e,g,i,n) and Fisher rat embryo cells (b,d,f,h) were grown to confluency in single wells (1 cm diameter) of 24-well cluster plates (Costar). Culture fluids were aspirated, and cells rinsed with 10 mM sodium phosphate pH 7.2, 100 mM NaCl and 5 mM MgCl_2 . A total of 0.05 ml of the same buffer containing 0.001 (j), 0.01 (k), 0.1 (l), 1.0 (m) or 10 (a,b,n) μg EGF, or partially purified 20,000- M_r preparations from P-100 columns of either TGF (1 mg; e, f) or SGF (1 mg; g,h), were added. Cultures were kept at 25°C for 1 min, then disrupted by incubation for 1 min at 4°C in 0.05 ml 10 mM sodium phosphate pH 7.2, 100 mM NaCl, 1% Triton X-100, 0.5% sodium deoxycholate and 5 mM MgCl_2 buffer. $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ (50 μCi) was added and incubation continued for a further 1 min. Reactions were stopped by addition of 0.05 ml of 0.65 M Tris-HCl pH 6.7, 1.0% SDS, 10% glycerol, 2.5% 2-mercaptoethanol and 0.1% bromophenol blue, heated for 2 min at 90°C and analysed by SDS-PAGE as described elsewhere⁹. Molecular weight standards include ^{14}C -labelled myosin (M_r 200,000), phosphorylase b (98,000), bovine serum albumin (69,000) and carbonic anhydrase (30,000).

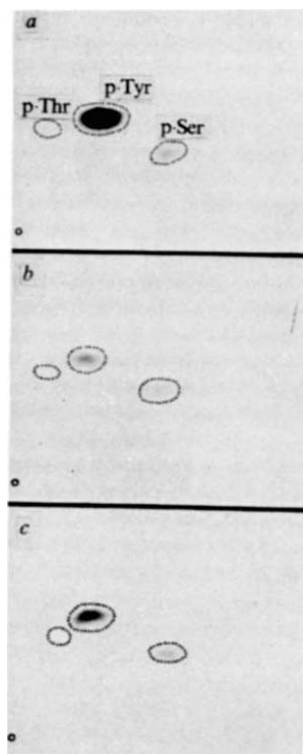


Fig. 2 Phosphoamino acid analysis of the 160 K EGF membrane receptors of A431 cells ^{32}P -labelled *in vitro* after exposure of cells to either mouse EGF or human TGF. A431 cells were incubated in medium containing mouse EGF ($1\ \mu\text{g ml}^{-1}$; a) or human TGF ($100\ \mu\text{g ml}^{-1}$; b), extracts prepared, and *in vitro* phosphorylation reactions performed as described in Fig. 1 legend. Snyder–Theilen FeSV P115 was immunoprecipitated by anti-FeLV from an extract of nonproductively transformed Fisher rat cells after labelling with $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ (c). $\gamma\text{-}^{32}\text{P}$ -labelled proteins were recovered from gel slices by incubation in TPCK–trypsin ($50\ \mu\text{g ml}^{-1}$ in $0.05\ \text{M}$ ammonium bicarbonate, pH 8.0; Worthington) for 6 h at 37°C . Supernatant fluids were filtered, lyophilized, washed by resuspension in H_2O and re-lyophilized. Tryptic peptides were hydrolysed in $6.0\ \text{M}$ HCl at 100°C for 1 h and two-dimensional phosphoamino acid determination performed as described elsewhere¹³. The positions of unlabelled standards including phosphothreonine (p-Thr), phosphotyrosine (p-Tyr) and phosphoserine (p-Ser) are indicated by dotted circles.

^{32}P -labelled P160 in extracts of either TGF(a)- or EGF(b)-treated cells contained phosphotyrosine as its major labelled phosphoamino acid, with smaller amounts of phosphothreonine and phosphoserine. Similar results were obtained for P160 isolated from extracts of SGF-treated cells. The relative proportions of ^{32}P -labelled phosphoamino acids in P160 after growth factor-activated phosphorylation were comparable with those found in Gardner FeSV P115 (Fig. 2c), a representative type C virus-encoded transforming protein with tyrosine-specific protein kinase activity⁹.

Although each of the growth factors examined here compete with EGF for binding to its receptor, only SGF and TGF phenotypically transform cells in culture. Analysis of the 160K EGF receptor in A431 cells by tryptic peptide mapping after either EGF- or human TGF-induced phosphorylation, revealed one major (no. 6) and seven minor phosphorylated peptides with corresponding positions, whether the tyrosine kinase was activated by EGF or by TGF (Fig. 3). On phosphoamino acid analysis of individual peptides, each was shown to contain tyrosine as its major phosphorylated species.

Cells transformed by several different RNA tumour viruses show an increase in the overall phosphotyrosine content of their proteins while those transformed by other RNA tumour viruses (mouse sarcoma, rat sarcoma)^{5,7–11,13,22} resemble the human tumour cell lines examined here in that their phosphotyrosine levels are not significantly greater than in untransformed cells. These differences may reflect the fact that viral transforming proteins encoded by avian sarcoma viruses^{5–8}, Gardner and

Snyder–Theilen FeSV^{9–11} and AbLV^{12,13} are expressed at relatively high levels and all possess intrinsic acceptor sites for tyrosine-specific protein kinase activities. Similarly, the increase in total cellular phosphotyrosine in A431 cells in response to EGF or TGF stimulation may be due to the exceptionally large number of EGF receptors ($2\text{--}3 \times 10^6$ per cell) in the A431 cell line. Thus, the inability to detect overall increases in phosphotyrosine in most spontaneously and chemically transformed cells does not argue against a pivotal role of tyrosine phosphorylation in the maintenance of transformation of such cells.

The phenotypic response of cells to both human and mouse cell-derived TGFs is strikingly different from their response to EGF, whereas A431 cells show similar phosphorylation responses to these two classes of growth factor. It is possible that, in addition to their binding to, and phosphorylation of, the 160 K major EGF membrane receptor, TGFs interact with an unidentified TGF-specific cell receptor. Alternatively, interaction of growth factors with, and resulting tyrosine phos-

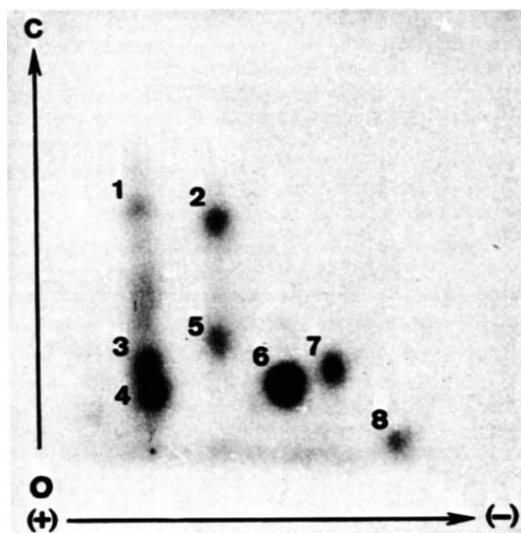


Fig. 3 Two-dimensional tryptic analysis of phosphopeptides derived from the 160 K EGF membrane receptor of A431 cells ^{32}P -labelled *in vitro* in the presence of human TGF. Extracts were prepared and *in vitro* phosphorylation reactions performed as described in Fig. 1 legend. ^{32}P -labelled proteins were recovered from gel slices by incubation in TPCK–trypsin as described in Fig. 2 legend, then incubated for 2 h at 4°C in $0.1\ \text{ml}$ of chilled performic acid (30% H_2O_2 and 90% formic acid (1:9) preincubated for 2 h at room temperature), diluted with $2\ \text{ml}$ H_2O , lyophilized, resuspended in electrophoresis buffer (acetic acid/formic acid/water, 15:5:80, pH 1.9) and spotted on to $10 \times 10\ \text{cm}$ TLC glass plates. Electrophoresis was for 30 min at $500\ \text{V}$ and chromatography was done in the second dimension in buffer containing butanol, pyridine, acetic acid and water (32:25:5:20). Radiolabelled tryptic peptides were visualized by autoradiography.

phorylation of, the 160 K EGF receptor could be part of the pathway leading to the transformed phenotype. Minor differences in the tryptic phosphopeptide maps of the P160 receptor labelled in response to EGF- and TGF-induced phosphorylation might reflect differences in the interaction of P160 with these growth factors.

The significance of protein kinases which have specificity for tyrosine acceptor sites both in RNA tumour virus-transformed cells and in response to growth regulatory factors is not fully understood. Recent reports of a series of cellular protein kinases, which are activated by phosphorylation of tyrosine acceptor sites and in turn have specificity for tyrosine residues, may be relevant²³. Activation of this pathway leads to phosphorylation of a tyrosine regulatory site(s) on the cellular ($\text{Na}^+ + \text{K}^+$)ATPase and seems to involve the cellular analogues of certain RNA tumour viruses²⁴. It remains to be determined whether phosphorylation of cellular receptors in response to TGFs directly leads to cell transformation by activation of a phosphorylation cascade.

This work was supported by NCI contract no. NOI-CO-75380.

Received 2 March; accepted 7 May 1981.

1. Todaro, G. J., Fryling, C. & De Larco, J. E. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5258-5262 (1980).
2. De Larco, J. E. & Todaro, G. J. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4001-4005 (1978).
3. Todaro, G. J. & De Larco, J. E. in *Control Mechanisms in Animal Cells: Specific Growth Factors* Vol. 1 (eds Jimenez de Asua, L., Levi-Montalcini, R., Shields, R. & Iacobelli, S.) 223-243 (Raven, New York, 1980).
4. Eckhart, W., Hutchinson, M. A. & Hunter, T. *Cell* **18**, 925-933 (1979).
5. Sefton, B. M., Hunter, T., Beemon, K. & Eckhart, W. *Cell* **20**, 807-816 (1980).
6. Collett, M. S., Purchio, A. F. & Erikson, R. L. *Nature* **285**, 167-169 (1980).
7. Feldman, R. A., Hanafusa, T. & Hanafusa, H. *Cell* **22**, 757-765 (1980).
8. Pawson, T. *et al.* *Cell* **22**, 767-775 (1980).
9. Reynolds, F. H. Jr, Van de Ven, W. J. M. & Stephenson, J. R. *J. biol. Chem.* **255**, 11040-11047 (1980).
10. Barbacid, M., Beemon, K. & Devare, S. G. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5158-5162 (1980).
11. Reynolds, F. H. Jr, Van de Ven, W. J. M., Blomberg, J. & Stephenson, J. R. *J. Virol.* **37**, 643-653 (1981).
12. Witte, O. N., Dasgupta, A. & Baltimore, D. *Nature* **283**, 826-831 (1980).
13. Blomberg, J., Reynolds, F. H. Jr, Van de Ven, W. J. M. & Stephenson, J. R. *Nature* **286**, 504-507 (1980).
14. Ushiro, H. & Cohen, S. *J. biol. Chem.* **255**, 8363-8365 (1980).
15. Fabricant, R. N., De Larco, J. E. & Todaro, G. J. *Proc. natn. Acad. Sci. U.S.A.* **74**, 565-569 (1977).
16. Haigler, H., Ash, J. F., Singer, S. J. & Cohen, S. *Proc. natn. Acad. Sci. U.S.A.* **75**, 3317-3321 (1978).
17. Blomberg, J., Van de Ven, W. J. M., Reynolds, F. H. Jr, Nalewaik, R. P. & Stephenson, J. R. *J. Virol.* **38**, 886-894 (1981).
18. Lee, L.-S. & Weinstein, I. B. *Science* **202**, 313-315 (1978).
19. Shoyab, M., De Larco, J. E. & Todaro, G. J. *Nature* **279**, 387-391 (1979).
20. Brown, K. D., Dicker, P. & Rozengurt, E. *Biochem. biophys. Res. Commun.* **86**, 1037 (1979).
21. Marquardt, H., DeLarco, J. E. & Todaro, G. J. (in preparation).
22. Reynolds, F. H. Jr, Van de Ven, W. J. M. & Stephenson, J. R. *J. Virol.* **36**, 374-386 (1980).
23. Spector, M., O'Neal, S. & Racker, E. *J. biol. chem.* **256**, 4219-4227 (1981).
24. Spector, M., Pepinsky, R. B., Vogt, V. M. & Racker, E. *Cell* (in the press).

Somatostatin modulates effects of angiotensin II in adrenal glomerulosa zone

Greti Aguilera, James P. Harwood* & Kevin J. Catt

Endocrinology and Reproduction Research Branch,
National Institute of Child Health and Human Development,
National Institutes of Health, Bethesda, Maryland 20205, USA

*Food and Drug Administration, Washington, DC 20204, USA

The octapeptide angiotensin II is a major regulator of the adrenal glomerulosa zone, acting both as an acute stimulus of aldosterone secretion and as a trophic hormone which increases steroidogenic enzymes and angiotensin II receptors in glomerulosa cells^{1,2}. Angiotensin II also mediates the adrenal effects of altered sodium balance, and is essential for the aldosterone response to sodium restriction^{3,4}. However, the adrenal effects of angiotensin II are attenuated during sodium loading, suggesting that other local or humoral factors modulate its actions on adrenal glomerulosa function^{5,6}. Somatostatin, the somatotropin release inhibiting factor of the hypothalamus, has been shown to inhibit the secretion and action of several pituitary^{7,8} and non-pituitary hormones⁹⁻¹³. Because somatostatin has been found in several non-neural tissues, and seems to act as a local regulator of endocrine function, we have now examined the possibility that it may also modulate the effects of angiotensin II in the adrenal glomerulosa cell. Our studies have shown that low concentrations of somatostatin specifically inhibit the production of angiotensin II-stimulated aldosterone, and that this action is mediated by specific, high-affinity receptors for somatostatin in the zona glomerulosa.

When incubated with collagenase-dispersed rat adrenal glomerulosa cells¹⁴, somatostatin markedly inhibited the ability of angiotensin II to stimulate aldosterone formation. As shown in Fig. 1, the maximal aldosterone response to angiotensin II was reduced by 43% in the presence of 1 nM somatostatin, and was abolished by 1 μ M somatostatin. In contrast, the aldosterone

responses to other stimuli, including potassium, ACTH and 8-bromo-cyclic AMP, were completely unaffected by somatostatin at concentrations of up to 1 μ M. Although low concentrations of somatostatin (10^{-10} – 10^{-9} M) had no effect on basal aldosterone production, higher concentrations ($>10^{-8}$ M) had a minor stimulatory effect on aldosterone production. The latter effect was inconstant and varied with the source and batch of the peptide, whereas the inhibitory effect of low concentrations of somatostatin was always evident. The inhibitory effects of somatostatin on the stimulation of aldosterone by increasing concentrations of angiotensin II are shown in Fig. 2. Without changing the basal aldosterone production, 10^{-10} and 10^{-9} M somatostatin reduced the maximal aldosterone responses to angiotensin II. Higher somatostatin concentrations (10^{-8} and 10^{-7} M) also decreased the maximal responses to angiotensin II, although basal aldosterone production was elevated due to the direct stimulatory action of the somatostatin preparation used in this experiment. Micromolar concentrations of somatostatin completely abolished aldosterone responses to angiotensin II concentrations up to 10^{-6} M.

To elucidate the mechanism by which somatostatin exerts its specific inhibitory effect on angiotensin II action, we analysed the interactions of synthetic somatostatin with adrenal receptors for angiotensin II and somatostatin. In ¹²⁵I-angiotensin II binding studies, several somatostatin preparations completely inhibited angiotensin II binding, but with varying half-maximum potencies (ID_{50}) of $4.2 \pm 1.0 \times 10^{-8}$ M, $3.8 \pm 2.5 \times 10^{-7}$ M and $5.1 \pm 1.2 \times 10^{-6}$ M for Beckman lots E0535 and B00926, and Peninsula peptides, respectively. These concentrations were similar to those required for stimulation of aldosterone production by the individual preparations, suggesting that interaction with the angiotensin receptor is responsible for their weak steroidogenic effects. This possibility was supported by the ability of the angiotensin II antagonist [Sar¹, Ala⁸]angiotensin II in three experiments (not shown) to inhibit the aldosterone responses to both 10^{-9} M angiotensin II and 10^{-7} M somatostatin, while the ACTH-stimulated aldosterone response was unchanged. The basis of the direct actions of certain somatostatin preparations on the angiotensin II receptor is not known, but such high-dose effects are not relevant to our main finding, that low concentrations of somatostatin consistently inhibit the aldosterone response to angiotensin II.

The specific inhibition of angiotensin-stimulated aldosterone responses by low somatostatin concentrations could not be explained by interaction with the angiotensin II receptor. Thus, nanomolar concentrations of somatostatin did not influence

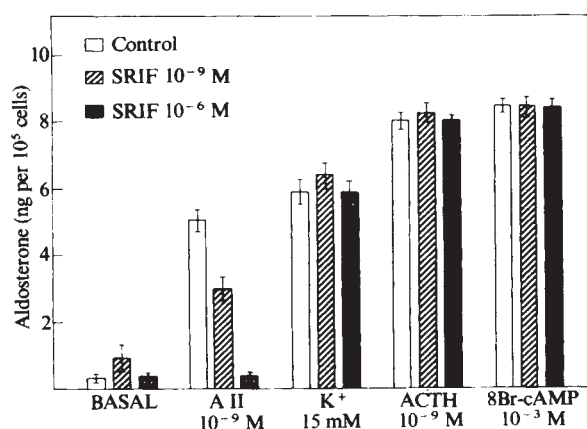


Fig. 1 Effect of somatostatin (Beckman, lot E0535) on basal and stimulated aldosterone production in collagenase-dispersed rat adrenal glomerulosa cells. Aliquots containing $1-2 \times 10^5$ cells were incubated for 2 h and aldosterone production was determined by direct radioimmunoassay of the incubation media¹⁴. Bars represent the mean and s.e. of three experiments. AII, angiotensin II; 8Br-cAMP, 8-bromo-cyclic AMP.

angiotensin II binding, yet significantly decreased the aldosterone responses to angiotensin II, suggesting that the peptide could act through specific receptors in the glomerulosa zone. In binding studies with ^{125}I -Tyr¹-somatostatin, specific high-affinity receptor sites for somatostatin were found in the adrenal capsule, which contains most of the zona glomerulosa layer (Fig. 3). In contrast, particulate fractions from the decapsulated adrenal displayed only 10% of the binding found in adrenal capsules (data not shown). In six experiments, the mean binding capacity for somatostatin was 329 ± 64 fmol per mg of adrenal capsular protein, and the association constant (K_a) was $0.9 \pm 0.3 \times 10^{10} \text{ M}^{-1}$. The binding of ^{125}I -Tyr¹-somatostatin to adrenal capsular particles was saturable and specific, with no displacement by angiotensin II, [D-Ala⁶]gonadotropin releasing hormone, vasopressin, oxytocin and substance P, at concentrations up to 10^{-5} M . The specificity of the somatostatin receptor sites has been confirmed in studies which demonstrate that a series of somatostatin analogues displace ^{125}I -Tyr¹-somatostatin from adrenal capsular membranes with potencies that correlate with their ability to inhibit angiotensin II-stimulated aldosterone¹⁵.

Somatostatin has been identified by immunofluorescence studies in specific cells of several tissues including thyroid⁹, pancreas^{11,13} and gastric mucosa^{12,13,16}, and local release of the peptide has been proposed to control the function of endocrine cells. Evidence for such a mechanism in the glomerulosa zone was sought by analysing the endogenous somatostatin content of the adrenal capsule. Tissues were homogenized in 2 M acetic acid, boiled for 5 min and centrifuged at 15,000g. The supernatant was lyophilized and the somatostatin content of the extract was determined by radioimmunoassay¹⁷. The immunoreactive somatostatin in extracts from adrenal capsules was $1,817 \pm 200$ pg per mg, compared with 146 ± 15 , 7.6 ± 1.3 and 10.3 ± 2 pg per mg in extracts of decapsulated adrenals, urinary bladder and aorta, respectively. The values observed in adrenal capsule were comparable with those present in the pancreas and gastrointestinal tract¹², and were much higher than could be attributed to peptide binding to the specific receptor sites of the zona glomerulosa.

These data demonstrate the presence of somatostatin receptors and well defined actions of the tetradecapeptide in the

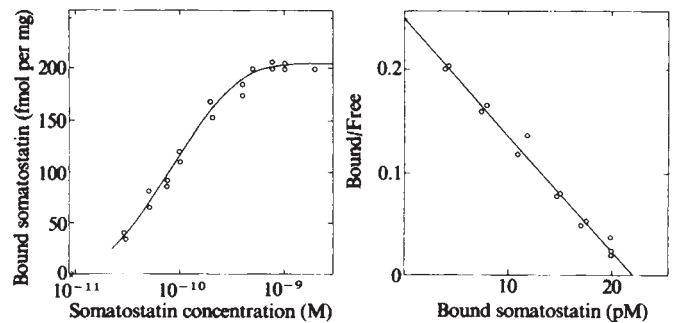


Fig. 3 Saturation curve and Scatchard plot of the binding of somatostatin (SRIF) to adrenal particles. Tyr¹-SRIF (Peninsula) was iodinated with ^{125}I by a modification of the technique of Hunter and Greenwood¹⁸, using prolonged incubation with low concentrations of chloramine T (1 μg). The reaction was terminated by addition of an excess of tyrosine, and the labelled peptide was purified by chromatography on a $0.7 \times 14 \text{ cm}$ Biogel P-2 column eluted with 0.01 M acetic acid. Specific activity of the tracer was 1,500–2,000 μCi per μg and maximum binding to excess receptors was 40–60%. Binding assays were performed by incubation of 50–100 μg of protein in 50 mM Tris-HCl buffer, pH 7.4, containing 10 mM MgCl_2 , 2 mM EGTA, 20 $\mu\text{g ml}^{-1}$ bacitracin, 50 $\mu\text{g ml}^{-1}$ Merthiolate and 1% bovine serum albumin in a total volume of 500 μl . Incubations were performed for 30 min at 20 °C, and bound ^{125}I -Tyr¹-SRIF was separated by filtration through Whatman GF/C glass fibre filters. Nonspecific binding was determined in the presence of 10^{-7} M unlabelled SRIF and this value (<1% of total radioactivity added) was subtracted from total binding to yield specific binding (8–15% of the total radioactivity added). Data points are the mean of duplicate incubations from a typical experiment.

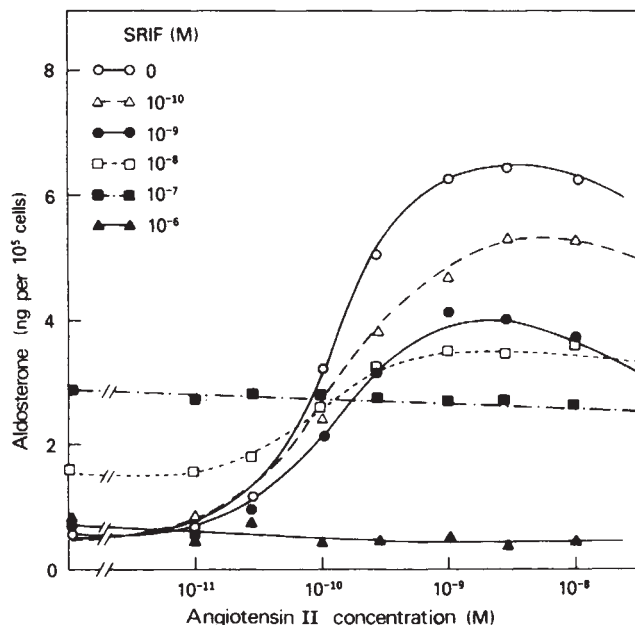


Fig. 2 Inhibition of aldosterone responses to angiotensin II (10^{-9} M) by somatostatin (SRIF) in collagenase-dispersed adrenal glomerulosa cells. Data points are the mean of duplicate incubations in one of three similar experiments.

adrenal glomerulosa zone. Somatostatin specifically inhibited the aldosterone responses to angiotensin II, and this effect was manifested at much lower concentrations of the peptide than those required to inhibit angiotensin II binding. Therefore, it is likely that somatostatin exerts its inhibitory effect on angiotensin II-induced aldosterone responses by interacting with specific high-affinity receptors and modulating the second messenger system that mediates the action of angiotensin II on aldosterone biosynthesis. The occurrence of high concentrations of immunoreactive somatostatin in the adrenal capsule suggests that, as in certain other tissues, somatostatin is locally secreted and could exert a regulatory action on target cell responses to endocrine stimulation. The highly specific inhibitory action of somatostatin on angiotensin II-stimulated aldosterone production could be relevant to the changing sensitivity of the adrenal glomerulosa zone to angiotensin II during the physiological control of aldosterone secretion.

We thank Dr S. Bathena for performing radioimmunoassay of somatostatin in tissue extracts.

Received 19 November 1980; accepted 12 May 1981.

- Hauger, R., Aguilera, G. & Catt, K. J. *Nature* **271**, 176–178 (1978).
- Aguilera, G., Menard, R. & Catt, K. J. *Endocrinology* **107**, 55–60 (1980).
- Aguilera, G. & Catt, K. J. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4057–4061 (1978).
- Aguilera, G., Schirar, A., Baukal, A. & Catt, K. J. *Circulation Res.* **46**, Suppl. 1, 118–127 (1980).
- Nichols, M. G. *et al.* *Endocrinology* **102**, 485–493 (1978).
- Cowley, A. W. & McCaa, R. E. *Circulation Res.* **39**, 788–797 (1976).
- Barzeau, P. *et al.* *Science* **179**, 77–79 (1973).
- Vale, W., Rivier, J., Brazeau, P. & Guilleman, R. *Endocrinology* **95**, 968–974 (1974).
- Kronheim, S., Berelowitz, M. & Pimstone, B. *Clin. Endocr.* **5**, 619–623 (1976).
- Bolaffi, L. L., Reichlin, S., Goodman, D. B. P. & Forrest, J. N. *Science* **210**, 644–646 (1980).
- Dubois, M. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1340–1343 (1975).
- Schlegel, W. *et al.* in *Hormonal Receptors in Digestive Tract Physiology* (eds Bonfils, S., Fromageot, P. & Rosselin, G.) 361 (North-Holland, Amsterdam, 1977).
- Guillemin, R. & Gerich, J. E. *A. Rev. Med.* **27**, 379–388 (1976).
- Douglas, J., Aguilera, G., Kondo, T. & Catt, K. J. *Endocrinology* **102**, 685–696 (1978).
- Aguilera, G., Parker, D. & Catt, K. J. *Endocrine Soc. 63rd a. Meet. abstr.* 16 (1981).
- Larsson, L., Golterman, N., DeMagistris, L., Rehfeld, J. F. & Schwartz, T. W. *Science* **205**, 1393–1395 (1979).
- Gerich, J., Greene, K., Hara, M., Rizzo, R. & Petton, G. *J. Lab. clin. Med.* **93**, 1009–1017 (1979).
- Hunter, W. M. & Greenwood, F. C. *Nature* **194**, 495–496 (1962).

Gonadotropin releasing hormone stimulates calmodulin redistribution in rat pituitary

P. Michael Conn*, James G. Chafouleas†, Deloris Rogers* & Anthony R. Means†

* Department of Pharmacology, Duke University Medical Center, Durham, North Carolina 27710, USA

† Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030 USA

Calcium (Ca^{2+}) seems to have an informational role in many tissues. In particular, it fulfills the requirements of a second messenger for gonadotropin releasing hormone (GnRH)-stimulated luteinizing hormone (LH) release from the pituitary gonadotrope (see ref. 1 for review). Very little is known about the effect of this ion on intracellular targets or the mechanism by which Ca^{2+} mobilization stimulates LH release. One intracellular target for Ca^{2+} is calmodulin, a ubiquitous intracellular Ca^{2+} receptor that has been shown to modulate many cellular functions, including cyclic nucleotide and glycogen metabolism, protein phosphorylation, microtubule assembly and disassembly, Ca^{2+} flux, and the activities of NAD kinase, tryptophan 5' monooxidase and phospholipase A_2 (see refs 2–5 for reviews). We have now used a specific and sensitive radioimmunoassay to determine the quantity and distribution of calmodulin in the gonadotrope before and during GnRH-stimulated LH release. The data indicate that GnRH stimulates redistribution of calmodulin from the cytosol to the plasma membrane and suggest that the molecule may have a role in the mechanism of stimulus-secretion coupling.

Figure 1 shows the distribution of calmodulin in the subcellular fractions (expressed as a percentage of the homogenate) at the indicated dose and time after GnRH injection. There is an initial rise in the percentage of calmodulin associated with the plasma membrane which appears concomitantly with the depletion of cytosolic calmodulin. The increase occurs with a time course similar to that of secretion of LH into the blood. As the calmodulin begins to disappear from the plasma membrane fraction, its level increases first in the mitochondrial/granule and microsomal fractions and finally in the cytosol. There is also a dose-response relationship between plasma membrane accumulation of calmodulin and its cytosolic depletion (Fig. 2). $\text{Des}^1\text{GnRH}^{2-10}$, which binds to the GnRH receptor with nearly 1,000-fold less affinity than GnRH itself⁶ and which had no efficacy in stimulating LH release *in vivo*, did not stimulate calmodulin redistribution (data not shown).

Calmodulin synthesis is constitutive in all systems examined (including the GnRH-stimulated pituitary)⁷. Accordingly, the redistribution reported here may indicate translocation between subcellular compartments. Accumulation of calmodulin at the plasma membrane at the expense of the cytosol might allow altered calmodulin control of regulatory functions at these loci. There is some evidence for such altered regulation by calmodulin in other systems. In the red blood cell, islet cell and adipocyte^{8–10} calmodulin-activated ATPases are located at the plasma membrane. In the case of the adipocyte, activation of Ca^{2+} ATPases seems to be hormonally regulated. Although calmodulin redistribution may be either a cause or a consequence of the secretory process, it does seem to have some role in secretion. Calmodulin has been found at the postsynaptic membrane, seems to mediate the Ca^{2+} effects on synaptic transmission^{11,12} and may, accordingly, have a role in the release of neurotransmitters. In addition, trifluoperazine (Stellazine),

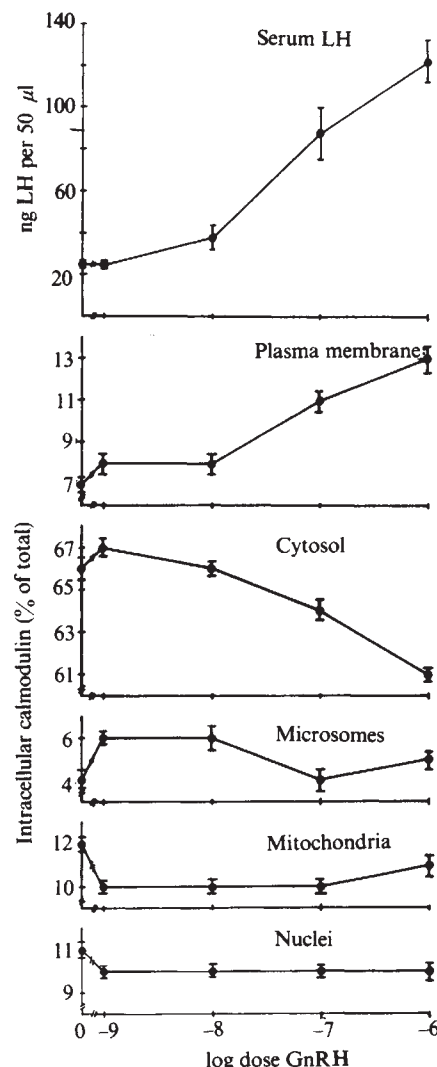


Fig. 1 GnRH concentration dependence of LH release and intracellular redistribution of calmodulin. GnRH (0.1–2 µg per rat, as indicated, obtained from the National Pituitary Agency) was administered by subcutaneous injection (0.1 ml in phosphate-buffered saline) to ovariectomized rats (Zivic-Miller, 5–7 weeks old, ovariectomized at 24 days). After 35 min the rats were killed by decapitation. Trunk blood was collected and serum LH determined by radioimmunoassay (materials from NIDDK, standard RP1, antibody LHS-5, LH for iodination I-5, prepared by Dr A. Parlow). The assay is sensitive to <1 ng per tube and cross-reacts <1% with follicle stimulating hormone. Pituitaries were removed and homogenized (1 ml per pituitary) with a Dounce all-glass homogenizer (five strokes A pestle, three strokes B pestle) in ice-cold 0.5 M sucrose, 50 mM Tris-HCl pH 7.4, 25 mM KCl, 2 mM MgCl_2 , 1 mM CaCl_2 (0.5 M sucrose-TKMC). The suspension was filtered through organza cloth which was further washed with 1 ml per pituitary of 0.5 M sucrose-TKMC. This filtered suspension was the homogenate fraction. After saving an aliquot for assay, the remainder of the homogenate was centrifuged at 1,000 g_{ave} for 10 min in a Beckman TJ-6 centrifuge. The pellet (P_1) was suspended by gentle homogenization (two strokes of the B pestle) in 2.0 M sucrose-TKMC and further centrifuged at 20,000 g_{ave} for 45 min (Sorvall SS-34 rotor). The pellet (P_2) and supernatant were designated nuclear and plasma membrane fractions, respectively. The supernatant of P_1 (S_1) was centrifuged at 12,000 g_{ave} for 20 min to collect the mitochondrial/granule fraction and the supernatant of this centrifugation further centrifuged at 100,000 g_{ave} (Beckman type 40 rotor) to collect the microplasmal (pellet) and cytosolic (supernatant) fractions. Aliquots of the homogenate and subcellular fractions were assayed for DNA, protein and LH or for marker enzymes (aldolase, glucose-6-phosphatase, malic dehydrogenase, NADH cytochrome c reductase and 5' nucleotidase) using previously described methods (refs 21, 22). These analyses indicated <10% cross-contamination between fractions. Other aliquots were stored at -70°C , then heated and assayed for calmodulin as previously reported²³. The calmodulin concentration in the homogenate was 0.56 ± 0.04 ng per µg protein. The values shown are means $\pm$ s.e.m.

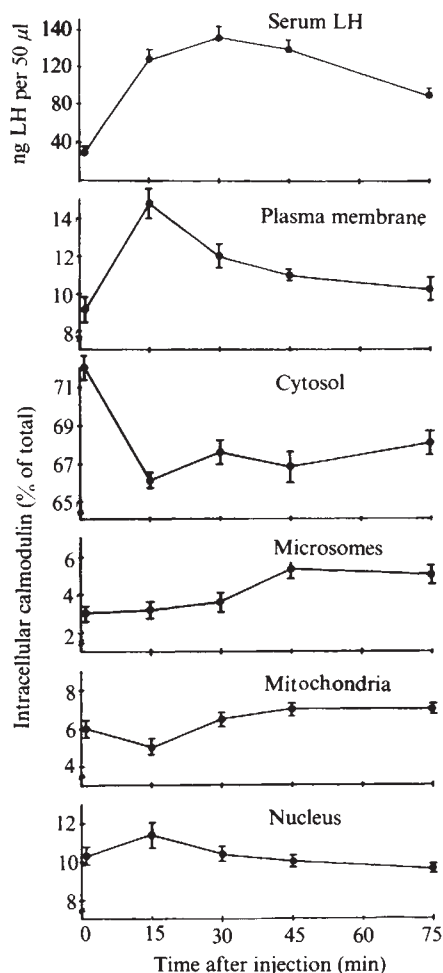


Fig. 2 Time course of LH release and subcellular calmodulin redistribution after a 2 µg injection of GnRH. Rats were killed at the indicated time. Subcellular fractions were prepared and assayed as described for Fig. 1.

which binds to calmodulin and inhibits some of its actions, also inhibits intestinal electrolyte secretion to the blood¹³.

Although internalization of GnRH does not seem to be required for the release process¹⁴ the gonadotrope responses to this releasing hormone include receptor patching, capping and ligand internalization¹⁵, as observed for many cell-surface receptor-mediated systems. Recruitment of the clathrin-coated vesicles to the plasma membrane can be observed as coated pits sensitive to phenothiazine calmodulin inhibitors¹⁶ and a recent study has shown that calmodulin is a constituent of the coated vesicle¹⁷. Thus, its appearance at the plasma membrane may be associated with the recruitment of coated pits involved in the internalization process. Insulin, which is also internalized¹⁸, is associated with the translocation of glucose transport activity from the microsomal or Golgi fractions to the plasma membrane^{19,20}. Thus, the phenomena of redistribution of new activities to the plasma membrane may be a generalized occurrence for plasma membrane receptor-mediated events and suggests a mechanism by which calmodulin regulated events could be effected in the absence of additional synthesis.

This work was supported by NIH HD07053 (A.R.M.), HD13220 (P.M.C.), RCDA NIH HD00337 (P.M.C.) and the Southern Regional Educational Board (P.M.C.).

Note added in the proof: Since acceptance of this article, it has been shown²⁴ that neurotropic agents are non-competitive antagonists of GnRH-stimulated LH release. The drug concentration needed to inhibit 50% of stimulated release correlated well with the ability to inhibit enzyme activation by calmodulin *in vitro*.

Received 29 January; accepted 23 April 1981.

- Conn, P. M. *et al. Endocr. Rev.* **2**, 174–185 (1981).
- Cheung, W. Y. *Science* **207**, 19–27 (1980).
- Klee, C. B., Crouch, T. H. & Richman, P. G. *Rev. Biochem.* **49**, 487–516 (1980).
- Means, A. R. & Dedman, J. R. *Nature* **285**, 73–77 (1980).
- Wang, J. H. & Waisman, D. M. *Curr. Topics cell. Regulation* **15**, 47–105 (1979).
- Marian, J., Cooper, R. L. & Conn, P. M. *Molec. Pharmac.* **19**, 399–405 (1981).
- Chafouleas, J. G., Conn, P. M., Dedman, J. R. & Means, A. R. *Endocrinology* **A106**, 289 (1980).
- Larsen, F. L. & Vincenzi, F. F. *Science* **204**, 306–308 (1979).
- Pershad Singh, H. A., McDaniel, M. L., Bry, C. G., Lacy, P. E. & McDonald, J. M. *Nature* **288**, 492–495 (1980).
- Pershad Singh, H. A. & McDonald, J. M. *Nature* **281**, 495–497 (1980).
- deLorenzo, R. J., Freedman, S. D., Yohe, W. B. & Maurer, S. C. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1838–1842 (1979).
- Grab, D. J., Berzins, K., Cohen, R. S. & Siekevitz, P. J. *J. biol. Chem.* **254**, 8690–8696 (1979).
- Ilundian, A. & Naftalin, R. J. *Nature* **279**, 446–468 (1979).
- Conn, P. M., Smith, R. G. & Rogers, D. J. *biol. Chem.* **256**, 1098–1100 (1981).
- Hazum, E., Cuatrecasas, P., Marian, J. & Conn, P. M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6692–6695 (1980).
- Salisbury, J. L., Condeelis, J. S. & Satir, P. J. *Cell Biol.* **87**, 132–141 (1980).
- Linden, C. D., Dedman, J. R., Chafouleas, J. G., Means, A. R. & Roth, T. F. *Proc. natn. Acad. Sci. U.S.A.* **78**, 308–312 (1981).
- Suzuki, K. & Kono, T. J. *biol. Chem.* **254**, 9786–9794 (1979).
- Suzuki, K. & Kono, T. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2542–2545 (1980).
- Cushman, S. W. & Wardzala, L. J. *J. biol. Chem.* **255**, 4758–4762 (1980).
- Wray, W., Conn, P. M. & Wray, V. P. in *Meth. Cell Biol.* **14**, 69–86 (1977).
- Schneider, W. in *Selected Data for Molecular Biology and Biochemistry*, K3–K14 (Chemical Rubber Press, 1970).
- Chafouleas, J. G., Dedman, J. R., Munjaal, R. P. & Means, A. R. *J. biol. Chem.* **254**, 10262–10267 (1979).
- Conn, P. M., Rogers, D. C. & Sheffield, T. *Endocrinol.* (in the press).

The expression-linked copy of surface antigen gene in *Trypanosoma* is probably the one transcribed

E. Pays, M. Lheureux & M. Steinert

Département de Biologie Moléculaire, Université libre de Bruxelles, 67, rue des Chevaux, 1640-Rhode St Genèse, Belgium

The antigenic specificity of the living trypanosome seems to be determined by the protein component of a unique glycoprotein species covering the whole surface of the parasite. During chronic infection, a single clone of trypanosomes may successively express a large repertoire of different variable antigen types (VATs). There are probably as many genes as variant-specific antigens (VSAs) (see refs 1–3 for reviews). The expression of the genes coding for the synthesis of these antigens is linked to genomic rearrangements involving duplication of the coding sequence and transposition of the additional copy^{4–6}. The regulation of the expression of the VSA genes is operated at the transcriptional level^{7–10}. It can thus be supposed that their transcription depends on the presence of the additional, transposed copy. We report here that this additional copy is in a chromatin configuration highly sensitive to pancreatic deoxyribonuclease, suggesting that it is the transcribed one.

The isolation of a cDNA clone coding for the AnTat 1.1 VSA of *Trypanosoma brucei brucei* (EATRO 1125, serodeme AnTar 1) has been described elsewhere⁹. Detailed *Pst*I restriction analysis and hybridization with probes prepared from different parts of this cloned AnTat 1.1 cDNA enabled us to show that several copies of the AnTat 1.1 sequence are present in distinct bands of the genomic DNA (ref. 6 and Fig. 1). Of these, only a 6.4-kilobase (kb) and a 2-kb band seem to contain all the information present in the cDNA⁶ and to share a common *Sst*I site 400 base pairs (bp) upstream of the coding sequence (Fig. 2 and manuscript in preparation); the other bands, which either contain incomplete sequences or show poor hybridization, are considered as imperfect copies of the the AnTat 1.1 gene⁶. The 3' site of all these *Pst*I fragments is the only *Pst*I site found in the cDNA, 340 bp from the 3' extremity (refs 6, 9 and Fig. 2). All these bands thus contain 5'-neighbouring genomic DNA as well as AnTat 1.1 information corresponding to the large (1,360-bp) 5' fragment of the cDNA. The remaining 3'-terminal *Pst*I sequence is not detected here, as our probe is the 550-bp *Hind*III–*Hind*III fragment (see Fig. 2). We have shown⁶ that the

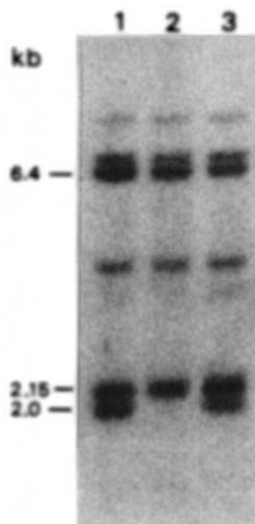


Fig. 1 Expression-linked copy of the AnTat 1.1 and AnTat 1.1b genes. The 32 P-labelled 550-bp *Hind*III–*Hind*III fragment of AnTat 1.1 cloned DNA was prepared as described previously^{6,9} and hybridized with Southern blots of *T. brucei brucei* nuclear DNA digested by *Pst*I. The DNA was extracted from bloodstream forms of the trypanosome expressing the VATs AnTat 1.1 (lane 1) or AnTat 1.1b (lane 3), or from procyclic forms derived from an AnTat 1.1 cloned population (lane 2) (see ref. 6 for details). The 2-kb band, which has clearly been lost in the procyclic trypanosomes, contains the additional, expression-linked copy of the VSA gene.

2-kb fragment is expression linked, being found only in the homologous AnTat 1.1 genomic DNA and not in DNA prepared from other variants of the same serodeme. The 6.4-kb *Pst*I fragment, present in the DNA of all the variants examined, is supposed to contain the basic copy of the AnTat 1.1 gene, that is, the one thought to be used as template to generate the expression-linked copy found in the 2-kb fragment⁶. It seems that the two copies are not contiguous, but rather that the additional copy has been transposed in a completely new genetic surrounding (Fig. 2). On the other hand, this additional copy is not extrachromosomal (data not shown). Finally, Hoeijmakers *et al.*⁴ have shown that the additional copy does not arise from DNA methylation of the target restriction site.

That the additional band is linked to transcription is already suggested by the fact that in all cases so far examined, a close correlation exists between the presence of the extra band and its corresponding mRNA (refs 4, 6, 9 and unpublished data). Moreover, we found that both the AnTat 1.1 mRNA and the 2-kb *Pst*I additional band (Fig. 1, lanes 1, 3) are present in two clone populations of different lineage expressing the same AnTat 1.1 VAT (ref. 6 and manuscript in preparation). On the other hand, in procyclic forms of the trypanosome, in which the expression of the VSA is turned down¹¹, the 2-kb AnTat 1.1 additional band was not detected (ref. 6 and Fig. 1, lane 2),

although these procyclic trypanosomes were directly derived, by *in vitro* cultivation, from an AnTat 1.1 clone.

The only firmly established difference between active and inactive chromatin in eukaryotes is the differential sensitivity of their DNA to DNase I^{12–17}. The transcribed sequences are indeed preferentially digested, probably because of the presence in active chromatin of either hyperacetylated histones^{18–26} or some high-mobility group proteins^{27–31}. In protozoa in particular, DNase I selectively releases high-mobility group proteins from *Tetrahymena* nuclei³⁰.

We prepared nuclei from trypanosomes expressing the VAT AnTat 1.1 and submitted them to mild digestion by DNase I. As clearly shown in Fig. 3, among the AnTat 1.1-specific *Pst*I bands, only the 2-kb additional sequence preferentially disappears

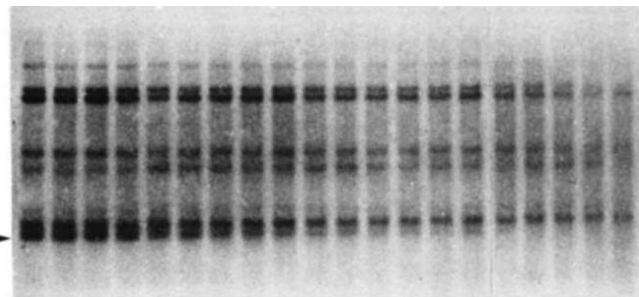
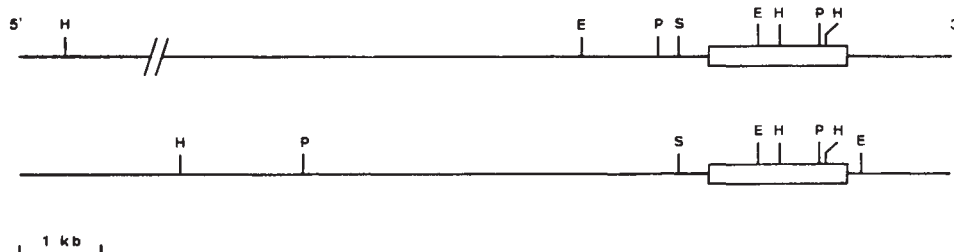


Fig. 3 Kinetics of the DNase I digestion of AnTat 1.1-specific sequences. Nuclei were purified from a clone population of trypanosomes expressing the VAT AnTat 1.1, by homogenization of the cells for 1 min at 23,000 r.p.m. in a Virtis 23 homogenizer, in 10 volumes of 0.32 M sucrose, 25 mM KCl, 3 mM MgCl₂, 20 mM Tris-HCl (pH 7.4) (medium A) at 4°C. After centrifugation at 1,500g for 10 min the pellet was resuspended in the same volume of medium A, but with 0.2% Nonidet P40 (Shell), and homogenized by hand in a Potter apparatus. After centrifugation as above, the nuclei were washed twice in 10 mM NaCl, 3 mM MgCl₂, 10 mM Tris-HCl (pH 7.4) (medium B), then incubated at 20°C for different periods as indicated, in 100-μl aliquots of medium B at 1 mg DNA ml⁻¹ with 80 ng ml⁻¹ DNase I (Worthington), previously freed from RNase by UMP agarose chromatography³⁴. The incubations were stopped by the addition of 5 volumes of 25 mM EDTA, 0.2% SDS and 250 μg ml⁻¹ proteinase K (Boehringer). After incubation for 15 h at 37°C, the DNA in each sample was extracted by phenol/chloroform (1/1), chloroform and ultimately by ether, incubated with 200 μg ml⁻¹ RNase A (previously freed from DNase by heating for 10 min in a boiling water bath) then re-extracted as above and precipitated in ethanol at -20°C. After centrifugation for 1 h at 16,000g, the DNA pellets were resuspended in 0.1 mM EDTA, 10 mM Tris-HCl (pH 7.5); 1 μg of each DNA sample was digested for 4 h at 37°C with 10 units of *Pst*I (Boehringer). The electrophoresis in 0.85% agarose gels, blotting on nitrocellulose filters, hybridization with a AnTat 1.1-specific cloned DNA probe and autoradiography were performed as described previously⁹. The different lanes contain, from left to right, a control DNA without DNase and samples incubated for 0, 5, 15, 30, 45, 60 s, 2, 3, 4, 5, 10, 15, 20, 30, 45, 60, 90, 120, 180 min with DNase. The arrow head indicates the additional band, containing the expression-linked copy of the VSA gene.

Fig. 2 Restriction maps of the basic (bottom) and additional (top) copies of the AnTat 1.1 gene and of their neighbourhood. Digestions with restriction endonucleases and hybridization of Southern blots of total genomic DNA digests with AnTat 1.1-specific probes were performed as described elsewhere⁶. E = *Eco*RI; H = *Hind*III; P = *Pst*I; S = *Sst*I. The 5' *Hind*III site to the left of the additional copy is located 19 kb from the next *Hind*III site in the AnTat 1.1 coding sequence. The putative extent of the coding sequence (1,700 bp, see ref. 9) is indicated by an open box.



after DNase I treatment. Because of their larger target size, the largest bands are affected on long DNase exposure. However, despite its larger size, the 6.4-kb *Pst*I fragment thought to contain the basic AnTat 1.1 sequence seems, among others, to be considerably less degraded than the expression-linked copy. If bands of similar size are compared, such as the 2.15- and the 2-kb fragments, the differential sensitivity to DNase I is particularly evident.

The chromatin of trypanosomes has not been extensively

Received 14 April; accepted 19 May 1981.

- Vickerman, K. *Nature* **273**, 613–617 (1978).
- Cross, G. A. M. *J. gen. Microbiol.* **113**, 1–11 (1979).
- Cross, G. A. M., Holder, A. A., Allen, G. & Boothroyd, J. C. *Am. J. trop. Med. Hyg.* **29**, 1027–1032 (1980).
- Hoeijmakers, J. H. J., Frasch, A. C. C., Bernards, A., Borst, P. & Cross, G. A. M. *Nature* **284**, 78–80 (1980).
- Borst, P. *et al. Am. J. trop. Med. Hyg.* **29**, 1033–1036 (1980).
- Pays, E., Van Meirvenne, N., Le Ray, D. & Steinert, M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2673–2677 (1981).
- Lheureux, M., Lheureux, M., Vervoort, T., Van Meirvenne, N. & Steinert, M. *Nucleic Acids Res.* **7**, 595–609 (1979).
- Hoeijmakers, J. H. J., Borst, P., Van den Burg, J., Weissmann, C. & Cross, G. A. M. *Gene* **8**, 391–417 (1980).
- Pays, E. *et al. Nucleic Acids Res.* **8**, 5965–5981 (1980).
- Agabian, N., Thomashow, L., Milhausen, M. & Stuart, K. *Am. J. trop. Med. Hyg.* **29**, 1043–1049 (1980).
- Barry, J. D. & Vickerman, K. *Expl. Parasit.* **48**, 313–324 (1979).
- Weintraub, H. & Groudine, M. *Science* **193**, 848–856 (1976).
- Garel, A., Zolan, M. & Axel, R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4867–4871 (1977).
- Panet, A. & Cedar, H. *Cell* **11**, 933–940 (1977).
- Bellard, M., Gannon, F. & Chambon, P. *Cold Spring Harb. Symp. quant. Biol.* **42**, 779–791 (1978).
- Miller, D. M., Turner, P., Nienhuis, A. W., Axelrod, D. E. & Gopalakrishnan, T. V. *Cell* **14**, 511–521 (1978).
- Zaslouff, M. & Camerini-Otero, R. D. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1907–1911 (1980).
- Vidali, G., Boffa, L. C., Bradbury, E. M. & Allfrey, V. G. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2239–2243 (1978).
- Sealy, L. & Chalkley, R. *Nucleic Acids Res.* **5**, 1863–1868 (1978).
- Simpson, R. T. *Cell* **13**, 691–699 (1978).
- Davie, J. R. & Candido, E. P. M. *Proc. natn. Acad. Sci. U.S.A.* **75**, 3574–3577 (1978).
- Oberhauser, H., Csordas, A., Puschendorf, B. & Grunicke, H. *Biochem. biophys. Res. Commun.* **84**, 110–116 (1978).
- Shewmaker, K. K., Cohen, B. N. & Wagner, T. E. *Biochem. biophys. Res. Commun.* **84**, 342–349 (1978).
- Mathis, D. J., Oudet, P., Wasylyk, B. & Chambon, P. *Nucleic Acids Res.* **5**, 3523–3547 (1978).
- Bonner, J., Wallace, R. B., Sargent, T. D., Murphy, R. F. & Dube, S. K. *Cold Spring Harb. Symp. quant. Biol.* **42**, 851–857 (1978).
- Levy, B., Watson, D. C. & Dixon, G. H. *Nucleic acids Res.* **6**, 259–273 (1979).
- Vidali, G., Boffa, L. C. & Allfrey, V. G. *Cell* **12**, 409–415 (1977).
- Weisbrod, S. & Weintraub, H. *Proc. natn. Acad. Sci. U.S.A.* **76**, 630–634 (1979).
- Gazit, B., Panet, A. & Cedar, H. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1787–1790 (1980).
- Hamana, K. & Zama, M. *Nucleic Acids Res.* **8**, 5275–5288 (1980).
- Weisbrod, S. & Weintraub, H. *Cell* **23**, 391–400 (1981).
- Borst, P. & Hoeijmakers, J. H. J. *Plasmid* **2**, 20–40 (1979).
- Astolfi Filho, S., Martins de Sa, C. & Gander, E. S. *Molec. biochem. Parasit.* **1**, 45–53 (1980).
- Brisson, O. & Chambon, P. *Analyt. Biochem.* **75**, 402–409 (1976).

Structure of mouse metallothionein-I gene and its mRNA

Niall Glanville, Diane M. Durnam
& Richard D. Palmiter

Howard Hughes Medical Institute Laboratory,
Department of Biochemistry, University of Washington,
Seattle, Washington 98195, USA

Metallothioneins are small cysteine-rich proteins that bind heavy metals such as zinc, cadmium, copper and mercury^{1,2}. Recent interest in these proteins has focused on the part they play in zinc metabolism and heavy metal detoxification¹. Our interest in metallothionein genes stems largely from the observations that these proteins are inducible by both heavy metals and glucocorticoid hormones^{1,3}. To explore the regulation of these genes, we have isolated cDNA and genomic clones corresponding to mouse metallothionein-I (MT-I)⁴, and have used them to show that both inducers act at the transcriptional level *in vivo* and in a wide variety of cell lines^{5–8}. We have also shown that the MT-I gene is amplified during selection for cadmium resistance⁵. To investigate the mechanisms of gene regulation, knowledge of the primary DNA sequence is necessary. Here we present the entire sequence of mouse MT-I gene along with ~300 bases of 5' flanking region that presumably includes promoter and regulatory sites. The 5' mRNA sequence, defined by S₁ nuclease mapping, was combined with sequences of the coding and 3' untranslated regions obtained previously⁴ to allow a computer prediction of the most stable secondary structure of MT-I mRNA.

The sequence of the mouse MT-I gene and its flanking regions is shown in Fig. 1a. Sequencing was done using the Maxam-Gilbert method⁹ with the strategy illustrated in Fig. 1b. Comparison of the genomic sequence with that of the cDNA clone⁴ shows that the coding region of the gene is interrupted by two introns as predicted by heteroduplex mapping⁴. The three exons of the gene are indicated by large capitals in Fig. 1a while the introns and flanking regions are shown by small capitals. The intron–exon junctions were deduced by comparing the genomic and cDNA sequences. Both introns obey the GT–AG rule originally proposed by Breathnach *et al.*¹⁰—they start with the dinucleotide GT and end with AG. In addition, both have a 3'

studied, but seems to be organized in typical nucleosomes^{32,33} and to be made of the same histones as other eukaryotes³³. It can thus be reasonably assumed that in trypanosome nuclei also, the transcribed sequences are preferentially digested by DNase I. It is concluded that the additional, expression-linked copy of the VSA gene, which is in a chromatin configuration highly sensitive to DNase I, is most probably the one which is transcribed.

This research received support from the Fonds de la Recherche Scientifique Médicale.

polypyrimidine tract similar to the polypyrimidine regions identified in many other introns¹¹. In many genes, introns lie between functional domains of the encoded protein^{12–14}. However, the 20 cysteine residues of metallothionein which coordinate seven metal atoms into two domains¹⁵ are distributed throughout the molecule (Fig. 1a), making it difficult to assign functional domains to the three exons.

The 5' end of exon 1 was deduced by S₁ nuclease mapping¹⁶. Total RNA isolated from cadmium-resistant Friend erythroleukemia cells⁵ was hybridized to a probe generated by cutting a plasmid carrying the genomic MT-I gene at its unique *Bgl*II site (position +64, Fig. 1a) and labelling the 5' ends with ³²P. After S₁ nuclease digestion, the size of the resistant fragments was analysed by comparing it with a sequencing ladder of the *Bgl*II–*Ava*II fragment (+64 to –180, Fig. 1a). The results, shown in lane 1 of Fig. 2, indicate that the most abundant DNA fragment generated by S₁ nuclease comigrates with the G labelled +1 in Fig. 1a. Control experiments in which either MT-I mRNA or S₁ nuclease was omitted showed no DNA bands migrating in this region (lanes 2 and 3, Fig. 2). The consensus sequence, Y–C–A–Y (in which the A is position +1; Y represents unspecified pyrimidine nucleoside), derived from comparisons of 20 mRNA start sites¹⁷ is located three nucleotides downstream from the G labelled +1 in Figs 1a and 2. Perhaps steric hindrance from the mRNA cap structure retarded S₁ nuclease digestion of our DNA probe down to the last base-paired nucleotide. We obtained the same result when the S₁ nuclease digestion was done at either 37, 50 or 60 °C. This analysis indicates that the 5' end of exon 1 and the mature mRNA lies 74 nucleotides from the coding region of MT-I mRNA.

Analysis of the region 5' to exon 1 shows that, like many other eukaryotic genes that are transcribed by RNA polymerase B, there is a Goldberg–Hogness (TATAAA) sequence (boxed in Fig. 1a) located between nucleotides –23 and –28. The sequence GC^TCAATCT, which is also common to various eukaryotic genes¹⁸, is not present in the 5' flanking region of the MT-I gene. The 5' region does, however, contain two large palindromic sequences centred at positions –55 and –103 (underlined in Fig. 1a). We are now generating mutants in this region to determine which areas are involved in the regulation of MT-I gene expression.

Localization of the mRNA start site allowed us to write the complete sequence of MT-I mRNA (Fig. 3). Computer

a

TGAGTTCTCG TAAACTCCAG AGCAGCGATA GGCCGTAATA TCGGGGAAAG CACTATAGGG
 -241
 ACATGATGTT CCACACGTCA CATGGTCTGT CCTATCCGAG CCAGTCGTGC CAAAGGGGCG
 -191
 GTCCCGCTGT GCACACTGGC GCTCCAGGGA GCTCTGCACT CCGCCCGAAA AGTGCGCTCG
 Bgl I
 GCTCTGCCAG GACCGGGGCG GCGTGACTAT GCGTGGGCTG GAGCAACCGC CTGCTGGGTG
 Set I
 -51
 CAAACGCTTT GGCGCGGAC TCGTCCAACG ACTATAAGA GGGCAGGCTG TCCTCTAAGC
 -41
 EXON 1
 1 10 20 30 40 50 60
 GTCAACACGA CTTCAACGTC CTGAGTACCT TCTCTCACT TACTCCGTAG CTCCAGCTTC
 ACCAGATCTC GGAATCGACC CCAACTGCTC CTGCTCCACC GGAAGACTC CCGATCCTTG
 Bgl II
 GCTTTTAGAA TACCAAGTTG GGACCGCAGA GCGGAATCCC CGAGTTGTAG AGGCTTGGCG
 -21
 GGAATAGGCA CCTTTAGTTG GCGATTCACT CCGGTTCTTT CTAGAATCCG CTCTTGCAAA
 Xba I
 AGCCTTCATT AGTTACGAGT ATTGTGCAAC GGGTCCTTTG GCGGGGTTGG GGTAGGATT
 -11
 TAGACGCGCA AATGTCGGGT TCCTGATCAC CCAATTAGTG GGGACATCTG GGTGAGTCC
 Bcl I
 CAGGCATTAC TAACTTACT GTGAATTGCT TGAATTAAGA AAGAGGTGAA GGACCTTTAT
 -1
 GTCTTGGGAC TCAAGACAT AATCCCTGAC TTAACCTGTG AGGAGAAAAG TGGGGCTAGG
 -1
 CTCCCTGCGAG CTCCGAGGAG GACTTAGTGA ACTGAGCCGG GACTCGTGGT GTTGGGCACT
 Pet I
 GCTGTAATGC TGCCTCCCTC ATGCTGTCTT CTTTCTCCTC CCAGGGCGGT CCGTGCATTG
 EXON 2
 1 10 20 30 40 50 60
 CACCAAGCTCC TGGCGCTGCA AGAAGTGAAG GTGCACCTCC TGCAAGAAGA GTGAGTTGGG
 T S S C A C C K N C S C C T S C K K
 ACACCTTGGG TGGCGGCTAA GGTAGGGGG GGGGAATCCC TACAAAAGT GCTCTGAGAA
 -1
 ATGCTCTTTG CTTCCCGGAG GCCATTGAT TGTCTCGGGG ACAGAACTAT ACAGAGAACT
 -1
 ATTTAAAAA ACCGAGGTCT TCTCTGTTGG GGACAGGAAG CAGAGGTCTT CAGCCAGGCT
 EXON 3
 1 10 20 30 40 50 60
 GCTCTTCTCT CTTTCTTCTA GGTCTGTGCT CCGTGTGCTC GGTGGGCTGC TCCAAATGTG
 A Q G C C V C K A A G C K C C C A C A
 CCCAGGGCTG TGTCTGCAAA GGCCTCGGGG ACAAGTGCACT GTGCTGTGCC TGAATGTGACG
 Set II
 AACAGCGCTG CCACCACGTG TAAATAGTAT CGGACCAACC CAGCGTCTTC CTATACAGTT
 -1
 CCACCTGTT TACTAAACCC CCGTTTCTA CCGAGTACGT TAATAATAAA AGCTGTTTGG
 -1
 AGTCTAAGTC TGGTTTCTTG GTGTGGTTTG GCAATAAGAA ACTGGGGTGA CTTGATAGTC
 -1
 TGGGGATCTG GTTTTGGACC CCCTCTGACC TTTACCTCCG CCCTCTGGCC CTCACAGAGG
 -1
 GGTAAATGCT TTGGGTAAG CCAAGCTATA TCCATAAGC TTTCTCATGG AAAACAGCTG
 Hind III
 Pvu II

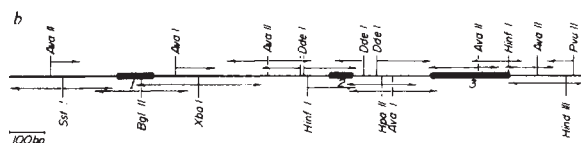


Fig. 1 The DNA sequence of the MT-I gene. **a**, Sequencing was done by the method of Maxam and Gilbert⁹. Exons are shown in large print; introns and the regions flanking the gene are shown in small print. The sequences of commercially available restriction enzymes that recognize 6-base pair (bp) sequences are underlined and named. Possible regulatory sites at the 5' end of the gene as discussed in the text are also underlined. The TATAAA sequence, initiation and termination codons are boxed. The MT-I protein sequence is indicated above the exon sequences using the one letter code: A, Ala; C, Cys; D, Asp; G, Gly; K, Lys; M, Met; N, Asn; P, Pro; Q, Gln; S, Ser; T, Thr; V, Val. **b**, Sequencing was done from the restriction sites indicated. DNA fragments were either ³²P-labelled at the 5' end using T4 polynucleotide kinase or at the 3' end using the large subunit of DNA polymerase I. Many of the sequence segments were determined on both DNA strands; all the sequence was analysed at least twice.



Fig. 2 Localization of the metallothionein-I mRNA start site by S₁ nuclease mapping. Lanes T, C, G and A: sequencing ladder⁹ from the BglII site (+64 in Fig. 1a). Labelling of the lanes from the sequencing gel has been transposed so that the mRNA sequence can be read directly. Lane 1, S₁ nuclease-resistant products obtained by digesting a hybrid between a DNA probe labelled at the BglII site and MT-I mRNA. A genomic clone was digested at the unique BglII site, treated with alkaline phosphatase, and labelled with [³²P]ATP using T4 polynucleotide kinase; the specific activity of the probe was 4 × 10⁶ c.p.m. per pmol of 5' ends. The probe was treated with proteinase K and phenol-chloroform extracted. Aliquots (4 × 10⁵ c.p.m.) were ethanol-precipitated with 140 μg of total RNA from cadmium-resistant Friend erythroleukemia cells⁵ (35 ng of MT-I mRNA). The precipitate was dissolved in 25 μl of 80% formamide, 0.4 M NaCl, 40 mM PIPES buffer (pH 6.5), 1 mM EDTA, overlaid with paraffin oil, heated at 80 °C for 5 min and incubated at 50 °C for 12 h. The samples were then diluted 10-fold with 10 mM Tris-HCl, pH 7.5 and ethanol-precipitated. The nucleic acids were dissolved in 25 μl of 10 mM NaCl and then 2.5 μl of buffer containing 3 M NaCl, 300 mM NaOAc, 30 mM ZnOAc, pH 4.5 was added. S₁ nuclease (2 μl, prepared by the method of Vogt²¹) was added and the sample incubated at 60 °C for 40 min. An aliquot (1 μl) was loaded onto the sequencing gel. Lane 2, conditions were the same as for lane 1, except that 140 μg of chicken calvaria RNA was hybridized with the probe. Lane 3, same as lane 1 except that no S₁ nuclease was added. A high-molecular weight band representing undigested ³²P-labelled DNA was visible on the gel but is not shown.

analysis¹⁹ of this sequence showed that various similar secondary structures were possible for MT-I mRNA. The most stable secondary structure predicted is shown in Fig. 3. This structure has an overall free energy of -87 kcal, suggesting that it is fairly stable. The model shows that most of the MT-I coding region, including the initiation codon (AUG), is contained within the duplexed region of the molecule, whereas the stop codon (UGA) is exposed in a loop. Note that the positions where RNA splicing occur (arrows, Fig. 3) are located very close to each other at the base of loop structures. It has been suggested that genes may evolve by addition or subtraction of exons²⁰. In order for this mechanism to yield a functional protein, the reading frame of the exons as well as a stable secondary and tertiary structure of both the protein and its mRNA must be preserved. Inspection of the secondary structure model of MT-I mRNA (Fig. 3) shows that the absence of exon 2 would have little effect on mRNA secondary structure. In addition, the reading frames of exons 1 and 3 are preserved without exon 2 (Fig. 1a). Thus, a plausible model for the evolution of MT-I gene might involve the primordial existence of an uninterrupted gene consisting of the nucleotides now represented by exons 1 and 3. This gene might then have been interrupted by an intron into which exon 2 was later inserted. Cell transformation techniques should allow us to reverse this process as a way of testing the hypothesis of exon addition.

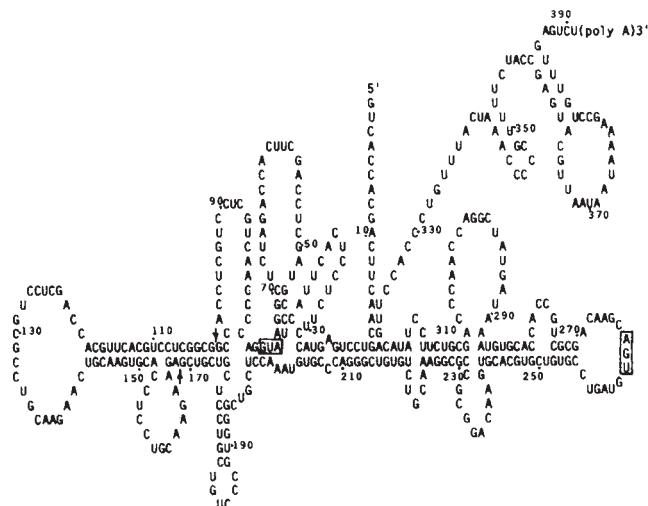


Fig. 3 Computer model of the secondary structure of MT-I mRNA. Analysis was done using the computer program developed by Pipas and McMahon¹⁹. The initiation codon (AUG) and the termination codon (UGA) are boxed. Splice points are indicated by arrows.

We thank Drs R. Gelinas, T. Gingeras and J. Millazzo for help with the computer analysis of MT-I mRNA secondary structure, and Ms Abby Dudley for secretarial assistance. This work was supported by NIH grant HD-09172. N.G. was a fellow of the Jane Coffin Childs Memorial Fund for Medical Research.

Received 22 April; accepted 18 May 1981.

- Kägi, H. R. & Nordberg, M. (eds) *Metallothionein* (Birkhäuser, Basel, 1979).
- Cherian, M. G. & Goyer, R. A. *Life Sci.* **23**, 1–10 (1978).
- Karin, M. & Herschmann, H. R. *Science* **204**, 176–177 (1979).
- Durnam, D. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 6511–6515 (1980).
- Beach, L. R. & Palmiter, R. D. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2110–2114 (1981).
- Mayo, K. E. & Palmiter, R. D. *J. biol. Chem.* **256**, 2621–2624 (1981).
- Durnam, D. M. & Palmiter, R. D. *J. biol. Chem.* **256**, 5712–5716 (1981).
- Hagar, L. J. & Palmiter, R. D. *Nature* **291**, 340–342 (1981).
- Maxam, A. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).
- Breathnach, R. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 4853–4857 (1978).
- Lewin, B. *Cell* **22**, 324–326 (1980).
- Sakano, H. *et al. Nature* **277**, 627–633 (1979).
- Lomedico, P. *et al. Cell* **18**, 545–558 (1979).
- Stein, J. P. *et al. Cell* **21**, 681–687 (1980).
- Otvos, J. D. & Armitage, I. M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7094–7098 (1980).
- Weaver, R. F. & Weissman, C. *Nucleic Acids Res.* **5**, 1175–1193 (1979).
- Corden, J. *et al. Science* **209**, 1406–1414 (1980).
- Benoist, C. *et al. Nucleic Acids Res.* **8**, 127–142 (1980).
- Pipas, J. M. & McMahon, J. E. *Proc. natn. Acad. Sci. U.S.A.* **72**, 2017–2021 (1975).
- Gilbert, W. *Nature* **271**, 501 (1978).
- Vogt, V. M. *Eur. J. Biochem.* **33**, 192–200 (1973).

Recombination between short DNA homologies causes tandem duplication

Thomas Edlund & Staffan Normark

Department of Microbiology, University of Umeå, S-901 87 Umeå, Sweden

The *ampC* gene of *Escherichia coli* K-12 codes for a β -lactamase which can hydrolyse the β -lactam ring of ampicillin^{1,2}. Ampicillin resistance is strictly related to *ampC* gene copy number thus we have been able to isolate ampicillin-resistant mutants carrying multiple *ampC* repeats^{3,4}. We have isolated on a plasmid a segment of chromosomal DNA carrying multiple *ampC* repeats, and compared the nucleotide sequence of the region joining repeat units to the sequence of the DNA segments that fused to create the joint. The fusion had occurred within a 12-base pair (bp) sequence of perfect homology. We suggest that recombination between randomly occurring short homologies (12–13-bp long), could be a general mechanism to generate tandem duplications in the size range of 10 kilobases (kb).

The properties of *E. coli* K-12 mutants carrying multiple copies of the *ampC* gene have been described elsewhere^{3,4}. We isolated amplified mutants that were derivatives of an *ampC* up-promotor mutant resistant to $\sim 20 \mu\text{g ml}^{-1}$ ampicillin⁵. By reciprocal recombination between a ColEI-*ampC* hybrid plasmid, pNU1 (ref. 4), and the chromosome of two *ampC*-amplified mutants, plasmid derivatives were isolated which carried multiple copies of the respective *ampC* duplication. One of these, derivative pNU8, carried multiple *ampC* genes of repeat size 9.8 kb. The repeat was found to have both its end points within the chromosomal DNA carried by plasmid pNU1 (ref. 4) thus plasmid pNU8 could be used as a source of DNA for the segments involved in the formation of the novel joint in the 9.8-kb repeat. The only difference between restriction fragment patterns obtained for pNU1 and pNU8 was the presence of an additional joint fragment in the latter. The restriction enzyme *PvuII* was used to generate a length of DNA small enough for sequence analysis but which also carried the joint fragment. Figure 1 shows the *PvuII* sites on plasmid pNU1. We concluded that the 9.8-kb repeat of pNU8 had one end point in the *PstI*₄-*PvuII*₅ fragment (fragment A) and the other in the *PvuII*₉-*PstI*₁ fragment (fragment B) of pNU1. Thus, the *PvuII* joint fragment (fragment C) of pNU8 consists of parts of fragments A and B.

The strategy for sequencing relevant stretches in fragments A, B and C is described in Fig. 2 legend. By comparing the sequences of these stretches (Fig. 3), it was found that the two DNA segments that had recombined contained a 12-bp long sequence of perfect homology, which was also present within the novel joint of fragment C. Thus, the 9.8-kb duplication had occurred by a recombination event at some point within these 12-bp sequences of fragments A and B.

E. coli strain G11a is usually resistant to $20 \mu\text{g ml}^{-1}$ of ampicillin. Mutants of G11a resistant to 40 or $60 \mu\text{g ml}^{-1}$ of

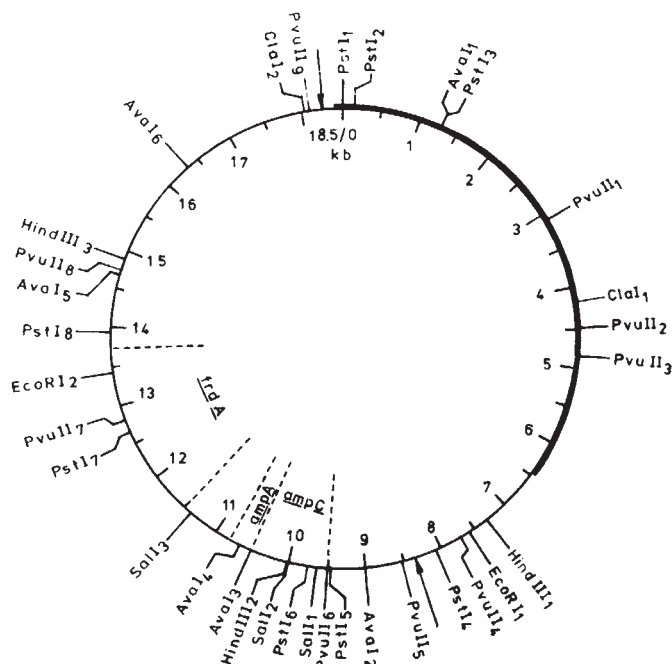


Fig. 1 Physical restriction fragment map of plasmid pNU1. The location of the *ClaI* and *PvuII* sites within pNU1 were determined by double digestions with these enzymes and the enzymes previously used to map pNU1 (ref. 4). The restriction endonuclease cleavage sites are numbered, starting from the *PstI*₁ site at the top. The thick-lined segment represents DNA from ColEI. Arrows indicate the end points for the 9.8-kb repeat unit. Locations are given for the *frdA* and *ampC* genes as well as the *ampA* control region.

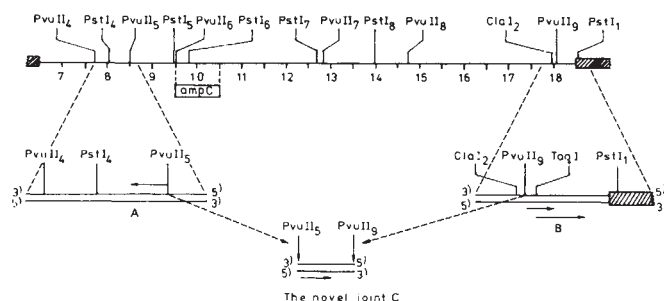


Fig. 2 Strategy for sequencing the novel joint of the 9.8-kb repeat, and the DNA segments involved in its formation. The numbered solid line at the top represents chromosomal DNA carried by plasmid pNU1. Hatched areas represent ColE1 DNA. The box indicates the position of the β -lactamase structural gene, *ampC*. The *TaqI* site within the *PvuII*₆–*PstI*₁ fragment was located by the mapping technique of Smith and Birnstiel⁸ using purified *ClaI*₂–*AvaI*₁ fragment of pNU1. For this purpose the fragment was end-labelled at the *ClaI* 5' terminus. The *PvuII*₅–*PvuII*₉ fragment carrying the novel joint was purified from plasmid pNU8 whereas the other fragments sequenced were from pNU1. Arrows indicate the direction and extent of sequencing runs, using the Maxam–Gilbert method⁹.

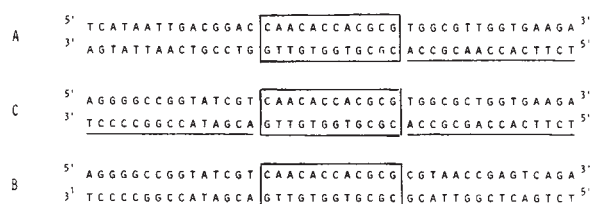


Fig. 3 Comparison of nucleotide sequences within fragments *PstI*₄–*PvuII*₅ (A) and *PvuII*₉–*PstI*₁ (B) of pNU1, and the *PvuII* 'joint' fragment (C) of pNU8. Boxed nucleotides represent a 12-bp perfect homology present in all three fragments. In fragment A this sequence was located ~90 bp from the *PvuII*₅ end, and in fragment B 230 bp from the *PvuII*₉ end. A recombination event within these 12 bp produced the novel joint.

ampicillin were isolated; this level of resistance corresponds to that shown by defined mutants carrying an *amp* duplication or triplication. By further selection for ampicillin resistance, amplifiable derivatives resistant to between 600 and 1,000 $\mu\text{g ml}^{-1}$ of ampicillin were obtained from individual *Amp*^r-40 or *Amp*^r-60 clones. Ten *amp*-amplified mutants were isolated in this way. The sizes and end points of their respective repeat units were determined by cleavage of chromosomal DNA with restriction endonucleases and the fragmentation patterns compared with that of pNU1. This ColE1 hybrid plasmid carries 12 kb of chromosomal DNA surrounding the *amp* gene. The

Table 1 Effect of size of homology on the length of DNA required to contain expected numbers of homologies

Homology size (bp)	No. of homologies				
	1	4	9	16	25
10	2*	3	4	5	6
11	3	5	7	9	11
12	5	9	13	17	21
13	9	17	25	33	41
14	17	33	49	65	81

Calculations were based on the fact that in a random sequence the number of homologies are proportional to the square of the length of the DNA segments expected to contain homologies.

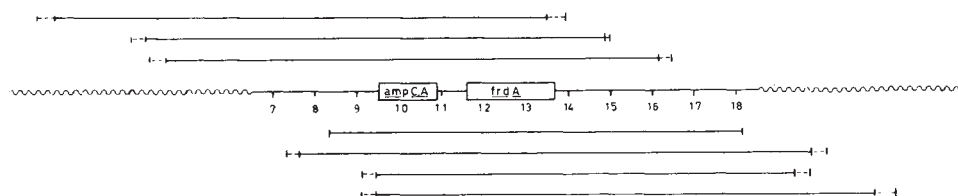
* The sizes of DNA fragments are given in kilobase pairs.

repeat copy number of the mutants was from 30 to 50 as judged from their ampicillin resistance, allowing restriction fragments of repeated chromosomal DNA larger than 400 bp to be detected. The sizes and end points of the respective repeats are given in Fig. 4. The end points could be accurately determined for seven amplified mutants; none of these had the same two end points. The sizes of the repeats varied between 9 and 13 kb. The remaining three mutants carried larger repeats (~18 kb). However, as their end points were situated outside the chromosomal DNA carried by pNU1, it was not possible to determine their exact position. However, we concluded that the end points of the ten different repeats were all located within 20 kb of chromosomal *amp* DNA.

A similar ampicillin selection procedure was performed on a *recA1* derivative of Gllal. Chromosomal DNA from 15 independently isolated *Amp*^r-300 derivatives was digested with several restriction endonucleases. No amplified fragments were observed in any of the mutants. These results suggest that the frequency of generation of *amp* duplications is decreased in a *recA1* background to such an extent that other types of ampicillin resistant mutants will predominate after the selection. On average, two homologous 12-bp sequences of any composition will be found within a random DNA sequence of 4 kb. This was deduced from the equation $x^2 = 4^n$, where n is the number of homologous base pairs and x is the number of base pairs which should contain a duplicate of any composition of n base pairs in length. This equation is an approximation which is valid if $x \gg n$.

In the *amp* system duplications may not occur within the *ampC* gene, which constitutes about 1 kb. Thus, on average, a duplicate of any 12-bp sequence will be found once in 5 kb. The general formula for this system will therefore be $(x - 1,000)^2 = 4^n$. The size of *amp* DNA required for different numbers of homologies with sizes ranging from 10 to 14 bp is given in Table 1; twenty-five 12-bp homologies and nine 13-bp homologies would be expected within a 21-kb and a 25-kb random sequence, respectively. In addition, our sequence and distribution data suggest that unequal recombination between any homologies close to 12 bp may be a general mechanism for

Fig. 4 Distribution of *amp* repeats in seven independent *Amp*^r mutants of *E. coli* K-12. The size and end points were determined by cleavage of chromosomal DNA with a series of restriction endonucleases. The amplified fragments were compared with the fragmentation patterns of plasmids pNU1 and pNU8. The numbered solid line indicates chromosomal DNA carried by pNU1; wavy lines represent surrounding chromosomal DNA. Horizontal bars above and below the genetic map indicate the extent of different *amp* duplications. Dashed lines indicate the extent to which the respective end point was mapped. The approximate locations for the *ampC* and the *frdA* genes are indicated by boxes. A represents the control sequence region for *ampC*.



generating tandem duplications. Our results also suggest that recombinations between such small homologies are promoted by the *recA* protein.

The novel joint described here was formed by recombination between DNA segments carrying a 12-bp perfect homology but with no other obvious homologies. This implies that the opportunity to form stable base pairing is an important feature of the mechanism that gives rise to tandem duplications. Promoter structures and possibly other controlling sequences which exist in many copies on the bacterial chromosome, could provide homologies for duplication, thus allowing catastrophic deletions. However, available sequence data on promoters indicate that they have been built up to contain recognition regions only a few base pairs long, interrupted by non-homologous stretches of nucleotides⁶. This strategy should prevent recombination and thus avoid deleterious deletion. In fact, it has been shown that intragenic deletions within the *lacI* gene can occur at repeated sequences as short as 5 bp⁷.

This work was supported by grants from the Swedish Natural Science Research Council (Dnr 3373-101) and the Swedish MRC (Dnr 5428).

Received 25 February; accepted 20 May 1981.

- Burman, L. G., Park, J. T., Lindström, E. B. & Boman, H. G. *J. Bact.* **116**, 123-130 (1973).
- Normark, S. & Burman, L. G. *J. Bact.* **132**, 1-7 (1977).
- Normark, S., Edlund, T., Grundström, T., Bergström, S. & Wolf-Watz, H. *J. Bact.* **132**, 912-922 (1977).
- Edlund, T., Grundström, T. & Normark, S. *Molec. gen. Genet.* **173**, 115-125 (1979).
- Jaurin, B., Grundström, T., Edlund, T. & Normark, S. *Nature* **290**, 221-225 (1981).
- Siebenlist, U., Simpson, R. B. & Gilbert, W. *Cell* **20**, 269-281 (1980).
- Farabaugh, P. J., Schmeissner, U., Hofer, M. & Miller, J. H. *J. molec. Biol.* **126**, 847-857 (1978).
- Smith, H. O. & Birnstiel, M. L. *Nucleic Acids Res.* **3**, 2387-2398 (1976).
- Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560-564 (1977).

Temperature-dependent ΔC_p^0 generated by a shift in equilibrium between macrostates of an enzyme

Harvey F. Fisher, Alan H. Colen
& Richard T. Medary

Laboratory of Molecular Biochemistry, Department of Biochemistry, University of Kansas School of Medicine and the Veterans Administration Medical Center, 4801 Linwood Blvd, Kansas City, Missouri 64128, USA

Substantial negative heat capacity changes (ΔC_p^0 's) have frequently been observed to accompany the formation of protein-ligand complexes^{1,2}. Glutamate dehydrogenase³ and horse liver alcohol dehydrogenase⁴, however, have been reported to form binary complexes with coenzyme with negligible ΔH^0 and only small ΔC_p^0 's. Although many intriguing mechanisms have been proposed to account for the observed phenomena, there is little direct experimental evidence available which might provide a basis for evaluating the contributions of ΔC_p^0 's of complex formation from the various mechanistic sources or even for distinguishing between them. However, if, as Eftink and Biltonen⁵ have suggested, a shift in equilibrium between macrostates contributes significantly to an observed ΔC_p^0 for a given reaction, it should be possible to characterize such a system by measuring the temperature dependence of the ΔC_p^0 . Despite this, few studies have determined ΔH^0 values at more than two temperatures. We have now measured the temperature dependence of the ΔH^0 (and, thereby, that of the ΔC_p^0) of the formation of an enzyme-reduced coenzyme complex in an attempt to provide such a basis and have found that the entire ΔC_p^0 of complex formation is accounted for by a temperature-induced shift of an equilibrium between the different forms of the free enzyme.

The temperature dependence of the calorimetrically measured ΔH^0 for the reaction written as



is shown in Fig. 1. Although the free enzyme is stable up to a temperature of $\sim 50^\circ\text{C}$ (ref. 6), the E-NADPH complex is markedly more heat-labile,⁷ and reliable measurements of the formation of this complex cannot be made at temperatures above 35°C using attainable flow rates. Precautions and controls described in Fig. 1 legend, however, assure us of the validity of measurements up to 35°C .

If the formation of the E-NADPH complex proceeded according to reaction (1) unaccompanied by a measurable ΔC_p^0 , the data of Fig. 1 would lie on a horizontal straight line, because $(\partial \Delta H^0 / \partial T)_p = \Delta C_p^0$. If the reaction proceeded with finite temperature-independent ΔC_p^0 , the data would lie on a straight line with a slope equal to ΔC_p^0 . It is apparent from Fig. 1 that the data correspond to neither simple case. The plot shows a marked curvature, the ΔC_p^0 approaching 0 at low temperatures and increasing to a value of $-450 \text{ cal K}^{-1} \text{ mol}^{-1}$ at higher temperatures. This behaviour can be explained on the basis of a temperature-sensitive equilibrium between 'implicit' macrostates of at least one reaction component⁸.

Consider a binding reaction which is thought to be expressed by reaction (1) but which, in fact, involves the following equilibria:



where

$$K_2 = \frac{[E]}{[E']} \quad \text{and} \quad K_1 = \frac{[E\text{-NADPH}]}{[E][\text{NADPH}]}$$

We assume that neither step involves a measurable ΔC_p . If step 2 is itself accompanied by a substantial ΔH^0 , and has a ΔG^0 such that $K_2 = 1$ within or near the experimental temperature range, the observed ' ΔH^0 ' for reaction (1) will vary from ΔH_1^0 , the value at temperatures at which the protein is entirely in the E form, to a value $(\Delta H_1^0 + \Delta H_2^0)$, at temperatures at which the protein is entirely in the E' form. By definition, the resulting temperature dependence of ΔH^0 will be interpreted as a ΔC_p^0 effect; and, if measurements are performed over a sufficient range of temperatures, the ΔC_p^0 itself will be found to be temperature dependent. This will be so even if, as we have assumed, both ΔH_1^0 and ΔH_2^0 are themselves temperature independent. We have portrayed the ligand as binding to only one form of the enzyme for ease in following the argument. In practice, no such distinction can be made from equilibrium measurements.

The dependence of ΔH^0 on temperature for reaction (2) is

$$\Delta H^0 = \Delta H_1^0 + \frac{\Delta H_2^0}{1 + K_2} \quad (3)$$

where $K_2 = \exp[\Delta H_2^0 (T - T_2) / RTT_2]$ and T_2 is that temperature at which $\Delta G_2^0 = -RT \ln K_2 = 0$.

The solid line in Fig. 1 shows the best fit of equation (3) to the data with the parameters $\Delta H_1^0 = 4.73 \pm 0.28 \text{ kcal mol}^{-1}$, $\Delta H_2^0 = -20.1 \pm 2 \text{ kcal mol}^{-1}$ and $T_2 = 43.6 \pm 1^\circ\text{C}$.

The magnitude of the apparent ΔC_p^0 generated by the effect of varying temperature on this implicit equilibrium is⁸

$$\left| \Delta C_p^0 \right| = \left| \left(\frac{\partial \Delta H^0}{\partial T} \right)_p \right| = \frac{K_2 (\Delta H_2^0)^2}{(1 + K_2)^2 RT^2} \quad (4)$$

For a reactant isomerization (such as reaction (1)), the sign of $\Delta C_p^{0'}$ is negative. For a reaction sequence involving a product isomerization step, however, the sign of $\Delta C_p^{0'}$ is positive.

In Fig. 2 we have plotted $\Delta C_p^{0'}$ as a function of temperature according to equation (4) using the best-fit parameters and algebraic sense of Fig. 1.

Because the sign of the $\Delta C_p^{0'}$ is negative, if there is a single (implicit) isomerization in the overall reaction, it must involve one of the reactants rather than one of the products. An attempt to fit the data to an equation similar to that for reaction (2) expanded to include an isomerization of the E-NADPH complex as well as one of E, indicated no contribution from that added feature. Thus, although there may be an isomerization of the enzyme-ligand complex, evidently such an equilibrium either is not poised in or near the experimental temperature range or must not involve a substantial $\Delta H^{0'}$. As reaction (2) is symmetrical in E and NADPH, an isomerization of the ligand could account for the data as well as an isomerization of E. While NADPH is known to exist as an equilibrium mixture of folded and unfolded forms, the T_2 for this process is below 20 °C and its $\Delta H^{0'}$ is too small to make an appreciable contribution to the observed process⁹. The attempted inclusion of finite intrinsic $\Delta C_p^{0'}$ terms for steps 1 and 2 of reaction (2) also failed to improve the fit of the equation to the data.

We conclude, therefore, that the observed $\Delta C_p^{0'}$ of the formation of the glutamate dehydrogenase-NADPH complex is fully accounted for by a poised isomerization of the free enzyme involving a very large enthalpy change that is largely compensated by an opposing large entropy change, the step in itself

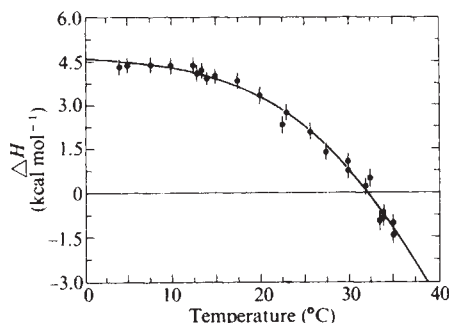


Fig. 1 The temperature dependence of $\Delta H^{0'}$ of the glutamate dehydrogenase-NADPH complex in 0.1 M phosphate buffer, pH 7.6. The reaction heats, q , were determined in a flow micro-calorimeter using apparatus, reagents and procedures as previously described³. The enzyme concentration was usually $\sim 90 \mu\text{M}$ (active sites). At each temperature, calorimetric measurements were made at saturating concentrations of NADPH starting at $20 \times K_D$ and extending to $200 \times K_D$. A plot of q versus [NADPH] was linear with a slope of $0 \pm 6 \times 10^{-4}$ kcal per mol NADPH in each case. As noted in the text, while the free enzyme is stable at 50 °C, its reduced coenzyme complex denatures rapidly above 37 °C. The following controls and precautions assure us that the data presented are unaffected by errors due to denaturation. (1) The solution emerging from the calorimeter was assayed for enzymatic activity. The activities so measured agreed with those of the initial solution to within 5% and showed no trend indicating denaturation. As the solution remains at an elevated temperature for some time after leaving the sensing chamber of the instrument, any time-dependent denaturation would be magnified several-fold by the time the solution actually emerged from the instrument. (2) The value of q was unchanged on raising the flow rate from $10 \mu\text{l s}^{-1}$ to $13 \mu\text{l s}^{-1}$. Had there been any significant time-dependent denaturation, the value of q would have differed because the two flow rates correspond to different residence times. The solid line is calculated from a nonlinear least-squares fit to reaction (3). The $\Delta H^{0'}$ value at 15 °C differs from that reported previously using a batch calorimeter. The batch value (and the very small $\Delta C_p^{0'}$ calculated from it) must now be presumed to be erroneous.

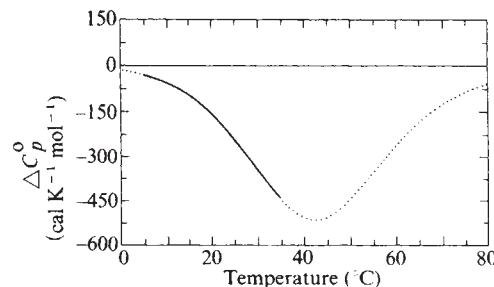


Fig. 2 The temperature dependence of $\Delta C_p^{0'}$ of the glutamate dehydrogenase-NADPH complex. The line is calculated from equation (4) using parameters from Fig. 1. The solid line portion of the curve indicates the region covered by experimental data.

involving no detectable $\Delta C_p^{0'}$ and H^+ transfer to or from the buffer not contributing significantly to the measured enthalpy change.

A similar mechanism, involving a temperature-induced shift between multiple forms of the free enzyme, is consistent with the recently reported large $\Delta C_p^{\ddagger}$ in the glutamate dehydrogenase-catalysed reaction⁸. We cannot say how general the phenomenon reported here may be, but we note that $\Delta C_p^{0'}$ s reported for complexes of other dehydrogenases at 20 °C are of a similar magnitude to that of glutamate dehydrogenase at its maximum near 40 °C. Thus, the differences between the $\Delta C_p^{0'}$ s of the various enzymes may simply reflect differences in the T_2 s of otherwise similar mechanisms.

Received 13 January; accepted 29 April 1981.

1. Sturtevant, J. M. *Proc. natn. Acad. Sci. U.S.A.* **74**, 2236-2248 (1977).
2. Biltonen, R. L. & Langerman, M. *Meth. Enzym.* **61**, 287 (1979).
3. Subramanian, S., Stickel, D. C. & Fisher, H. F. *J. biol. Chem.* **250**, 5885-5889 (1975).
4. Subramanian, S. *Biophys. Chem.* **7**, 375-378 (1978).
5. Eftink, M. & Biltonen, R. L. in *Biological Microcalorimetry* (ed. Beezer, A. E.) 396 (Academic, London, 1980).
6. Olsen, J. R. & Anfinsen, C. B. *J. biol. Chem.* **202**, 841-856 (1952).
7. Grisolia, S., Quijada, C. L. & Fernandez, M. *Biochim. biophys. Acta* **81**, 61 (1954).
8. Colen, A. H., Medary, R. T. & Fisher, H. F. *Biopolymers* **20**, 879-889 (1981).
9. Catterall, W. A., Hollis, D. P. & Walters, C. F. *Biochemistry* **8**, 4032-4036 (1969).

Corrigendum

In the letter 'Competition relatedness and efficiency' by J. R. W. Harris, *Nature* **292**, 54-55, the parameter x in equation (2) should be replaced by z .

Errata

In the article 'Seismotectonics of the El Asnam earthquake' by M. Ouyed *et al.*, *Nature* **292**, 26-31, the legend to Fig. 6 was reproduced incorrectly in some copies of *Nature*. It should read: *a*, The 1.30-m left lateral slip measured on a ploughed field at Zebadja near the centre of the fault. *b*, Southern segment of the fault. Typical pressure ridges showing the uplift of the north-western block (to the left), and collapse due to gravity. Extension cracks follow the compression axis.

In the news item 'Yugoslavian nuclear power: Locals late reacting', *Nature* **291**, 446, the name of the head of the Croatian Commission for the Human Environment was given incorrectly. It should read Marko Branica.

MATTERS ARISING

Grand unification magnetic monopoles inside the Earth

CARRIGAN¹ has suggested that the equilibrium between the magnetic and gravitational forces would lead to these monopoles being trapped in two positions on the magnetic axis, one each side of the centre, depending on their polarity. However, his discussion is incorrect in several respects. First, in the Earth's core, the gravitational field is roughly proportional to radius r . If the magnetic field were due to a central dipole it would be proportional to r^{-3} , and there would be (radial) equilibrium on the axis at $\sim r = 0.18 R_e$, within the fluid core, as he states. But the magnetic field is far from being dipolar, and within the core the dipole and non-dipole fields are comparable in magnitude; even at the core surface there are already ~ 14 positions² where the field is radial, not just the two he assumed.

Second, in fact, the magnetic field is produced by electric currents in the core; reasonable current distributions give a dipole radial component of field which increases only very slowly with depth. This would give an 'equilibrium' radius of $\sim 0.007 R_e$, only 50 km from the centre, and well within the solid inner core. Third, his 'equilibrium' calculations have apparently considered only radial forces, and for his dipole 'equilibrium' the monopole is unstable to infinitesimal tangential displacements. Stable equilibrium would require complicated fields with a most unlikely coincidence of magnitude and topology.

Finally, as Carrigan points out, in the core there are probably toroidal fields much larger than the poloidal dipole fields. However, these toroidal fields have no radial component, and will be largely latitudinal, so will not "result in large axial fields". What they will do is make it even less likely that there will be any stable positions of purely radial field. The inner core also will have large toroidal fields³.

It does seem most unlikely that there will be any significant trapping of monopoles by the mechanism Carrigan suggests.

F. J. LOWES

School of Physics, The University,
Newcastle upon Tyne, NE1 7RU, UK

1. Carrigan, R. A. Jr *Nature* **288**, 348–350 (1980).

2. Barracough, D. R. & Malin, S. R. C. *Geophys. J. R. astr. Soc.* **65**, 467–473 (1981).

3. Gubbins, D. J. *geophys. Res.* (in the press).

CARRIGAN REPLIES—Because of the limitations of space I could not emphasize the model nature of the calculation. However, I did point out that the dipole

model was not a proper picture of the interior of the Earth. I did not say a monopole of one polarity would occupy a single region of equilibrium but rather that GUMMs of one polarity would interchange with those of the other polarity. I did not touch on one aspect of the situation, the fact that GUMMs would probably not move rapidly through solid material in the interior of the Earth. If this is so it represents a third effective force. Static or dynamic equilibrium would be a result of all three effective forces.

In view of the limits of the picture—little is known about GUMM energy loss processes, there is incomplete knowledge of the interior of the Earth and little understanding of the accretion mechanism—it did not seem appropriate to develop a more comprehensive picture. The main point I wished to make was that even a relatively few GUMMs of opposite polarities could contribute a spectacular amount of energy.

RICHARD A. CARRIGAN JR

Fermi National Accelerator Laboratory,
PO Box 500, Batavia,
Illinois 60510, USA

Estimated speed of a giant bipedal dinosaur

FROM observations of locomotion in diverse living animals, including ostrich, bird, man, horse and elephant, Alexander¹ obtained a relationship between speed, stride length (λ) and body size (expressed by h , height at the hip). This relationship seems to hold true, at least in general terms, for large and small animals, both bipeds and quadrupeds, at gaits from walk ($\lambda/h < 2.0$) to run ($\lambda/h > 2.0$). To estimate the speeds of certain dinosaurs Alexander applied this relationship to their tracks, where λ could be measured directly and h could be estimated from the size of the footprints. He found¹ that the estimated speeds were rather low—between 1.0 and 3.6 m s⁻¹ (3.6 and 13.1 km h⁻¹).

Alexander's method has since been applied to various other dinosaur tracks^{2–5}, but few of these have yielded speed estimates greater than about 4.2 m s⁻¹ (15.0 km h⁻¹). A notable exception was provided by Russell and Béland², who calculated that a short section of trackway in the Mesaverde Formation (Cretaceous) of Colorado⁶ had been made by a dinosaur running at a speed of 7.54 m s⁻¹ (27.1 km h⁻¹). The Colorado dinosaur was apparently a giant bipedal ornithischian (probably an ornithomimid of the family Hadrosauridae), with h being

~ 3.44 m and a live body weight of ~ 11 tonnes². Russell and Béland based their calculations on data published by Brown⁶, whose account of the Colorado trackway is rather ambiguous; in one place Brown referred⁶ to a "fifteen foot stride", while in another he mentioned that the dinosaur "stepped 15 feet 2 inches". A stride is commonly understood to comprise two consecutive steps, or paces⁷, though the terms 'stride' and 'step' have sometimes been used synonymously. Evidently Brown was referring to a step (or pace), for the distance of 15 feet was measured between two footprints which he identified as those of left and right feet. Russell and Béland apparently followed Brown's interpretation: they presumably took the figure of 15 feet to represent a step (left footprint to right footprint) and doubled it to obtain the figure of 9.25 m which they cite for stride length².

I have examined photographs of the Colorado footprints (supplied by Dr E. S. Gaffney), and I suspect that Brown's interpretation of them may be wrong (see Fig. 1). The two footprints identified by Brown as left and right are very similar,

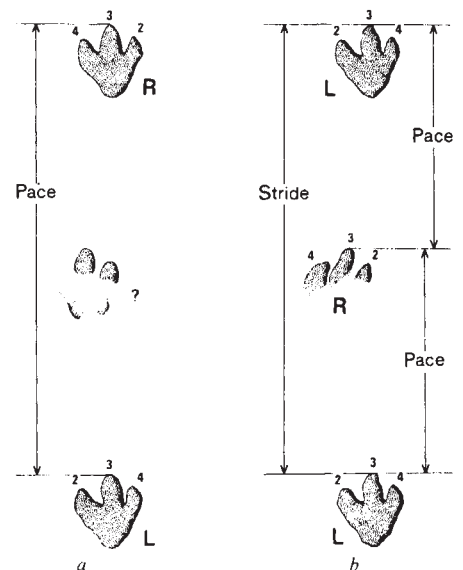


Fig. 1 Diagram illustrating two interpretations of ornithomimid dinosaur trackway from the Cretaceous of Colorado (American Museum of Natural History, no. 3650). Each footprint has the digits numbered and is identified as left (L) or right (R). The footprints are represented by natural casts on the under-surface of a sandstone bed, so that left and right appear transposed. *a*, Brown's interpretation⁶, indicating a pace of ~ 15 feet (4.57 m); the surface feature marked ? is not regarded as part of the trackway. *b*, Revised interpretation, indicating a pace of ~ 7.5 feet (2.29 m).

and they may well be prints of the same (left) foot. Both prints seem to show the same pattern of asymmetry—weak outward curvature of digit 3, and an angle of divarication between digits 2 and 3 which is slightly greater than that between digits 3 and 4. Mid-way between these two footprints is a poorly defined surface feature which was regarded by Brown⁶ as an amalgam of two footprints from other dinosaur tracks. This feature is probably an incomplete print of the ornithopod's right foot; it is about the same width as the other two footprints and shows traces of three digits with an appropriate pattern of divarication. The three footprints have a pattern of digit spacing (digit 2 slightly more divergent than digit 4) which is characteristic of many ornithopod footprints⁵ and which is consistent with their identification as left, right and left. This interpretation also reveals distinct positive (medial) rotation of the footprints, an arrangement which is common in the trackways of other bipedal dinosaurs (see, for example, refs 8 and 9); no such footprint rotation is evident in Brown's interpretation.

According to the revised interpretation (Fig. 1b) Brown's measurement of 15 feet would actually represent a stride (between successive prints of the same foot). Using Alexander's method¹, it may then be calculated that the Colorado ornithopod was walking (with λ/h of 1.34) at a speed of 2.4 m s^{-1} (8.5 km h^{-1}). The highest speeds so far estimated from dinosaur tracks are in the region of $3.6\text{--}4.3 \text{ m s}^{-1}$ ($13.0\text{--}15.5 \text{ km h}^{-1}$), and these are attributed to small bipedal runners with $h < 1 \text{ m}$ (refs 1,5). There are no reports of tracks indicating that larger bipedal dinosaurs were capable of achieving a running gait (with $\lambda/h > 2.0$).

I thank Dr E. S. Gaffney (American Museum of Natural History) for providing a description and illustrations of the Colorado footprints.

RICHARD A. THULBORN

Department of Zoology,
University of Queensland,
St Lucia, Queensland 4067,
Australia

1. Alexander, R. McN. *Nature* **261**, 129–130 (1976).
2. Russell, D. A. & Béland, P. *Nature* **264**, 486 (1976).
3. Tucker, M. E. & Burchette, T. P. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **22**, 195–208 (1977).
4. Coombs, W. P. Q. *Rev. Biol.* **53**, 393–418 (1978).
5. Thulborn, R. A. & Wade, M. *Lethaia* **12**, 275–279 (1979).
6. Brown, B. *Nat. Hist.* **41**, 190–202, 235 (1938).
7. Sarjeant, W. A. S. in *The Study of Trace Fossils* (ed. Frey, R. W.) 283–324 (Springer, New York, 1975).
8. Currie, P. J. & Sarjeant, W. A. S. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **28**, 103–115 (1979).
9. Newman, B. H. *Biol. J. Linn. Soc.* **2**, 119–123 (1970).

RUSSELL REPLIES—Brown clearly uses 'stride' in the sense of step in his description¹: "The tracks... (show) the right and left foot in normal stride where the giant

had stepped 15 feet". He also describes a third footprint, made when the animal had stepped on a firmer substrate. The steps defined by the three footprints would presumably have been of comparable length or Brown would not have referred them to the same trackway. In view of the symmetry of hadrosaur footprints² and Brown's presence at the site where the trackway was excavated, prudence suggests that his interpretation remains as plausible as Thulborn's proposed alternative.

DALE A. RUSSELL

Paleobiology Division,
National Museum of
Natural Sciences,
Ottawa, Canada K1A 0M8

1. Brown, B. *Nat. Hist.* **41**, 190–202, 235 (1938).
2. Langston, W. *Nat. Hist. Pap. natn. Mus. Can.* **4**, 1–9 (1960).

Inhibition of haemoglobin S gelation and water structural effects

BENESCH AND BENESCH¹ recently proposed an interesting hypothesis concerning the mechanism of action of various hydrophobic molecules which inhibit the gelation of sickle cell haemoglobin (HbS). They suggested that this hydrophobic class of sickling inhibitors may act on water structure rather than interact directly with HbS. The physical explanation provided by Benesch and Benesch¹ for the 'melting' of the HbS gel at low temperatures seems very plausible. They suggest that the creation of more polyhedral void spaces in water at lower temperatures and the consequent coalescence of 'filled' (with nonpolar moieties) and 'empty' polyhedral cages is responsible for weakening the hydrophobic forces. However, they extend this reasoning to suggest that the inhibition of polymerization of HbS by a variety of compounds containing hydrophobic residues is also "brought about by competition between the water polyhedra filled with small hydrophobic molecules and those attached to the protein, rather than a direct interaction between the hydrophobic residues and the protein which should be stereospecific"¹. The two situations are not physically analogous.

The clumping of 'empty' and 'filled' polyhedral cages leads to the relief of hydrophobic interactions while the clumping of filled cages with other filled cages produces hydrophobic interactions². This will apply equally whether the cages are filled with nonpolar segments from the protein or the small hy-

drophobic molecules. When the latter are added to a solution containing the HbS polymer, some existing water polyhedral void spaces will be further occupied and the resulting encounter of cages filled with small hydrophobic molecules with those filled with nonpolar side chains of the protein should provide the stimulus for a direct (hydrophobic) interaction of the small molecule with the haemoglobin nonpolar side chains, in competition with the side chain-side chain interaction. Gel melting can occur, in such cases, by a substitution of protein-protein hydrophobic interaction with protein-small molecule hydrophobic interaction. This interaction may be weak or strong, depending on the nature of the molecule and the side chain. If, in addition to this hydrophobic interaction, ionic or hydrogen bonding is possible, the direct interaction will be further strengthened. Such a possibility has been postulated for a molecule like benzyl alcohol³, although there has been no direct proof of this.

The attempts to increase the effectiveness of hydrophobic antisickling agents by structural modifications have failed so far, not because the agents are incapable of direct interaction with the protein but because of stronger macromolecule-macromolecule interactions in the polymer.

S. SUBRAMANIAN

Laboratory of Nutrition and
Endocrinology,
National Institute of Arthritis, Diabetes,
Digestive and Kidney Diseases,
National Institutes of Health, Bethesda,
Maryland 20205, USA

1. Benesch, R. & Benesch, R. E. *Nature* **289**, 637 (1981).
2. Stillinger, F. H. *Science* **209**, 451 (1980).
3. Ross, P. D. & Subramanian, S. *Biochem. biophys. Res. Commun.* **77**, 1217 (1977).

Matters Arising

Matters Arising is meant as a vehicle for comment and discussion about papers that appear in *Nature*. The originator of a Matters Arising contribution should initially send his manuscript to the author of the original paper and both parties should, wherever possible, agree on what is to be submitted. Neither contribution nor reply (if one is necessary) should be longer than 500 words and the briefest of replies, to the effect that a point is taken, should be considered.

BOOK REVIEWS

Food aid, self-sufficiency and the new technology

Barbara Huddleston

IN *Seeds of Plenty, Seeds of Want*, Andrew Pearse sets the stage for consideration of a number of new directions for food and agriculture policy in the 1980s. What to do about food aid is one of them. The subject is currently embroiled in controversy, much of it created by the uneven income distribution effects of Green Revolution technology, which food aid has more often than not reinforced.

How did food aid come to be what it is, and what changes can we expect in its use as an instrument of foreign policy and consequent impact on developing Third World economies? For answers to these questions, the interested reader should surely not miss Mitchel Wallerstein's new book, *Food for War — Food for Peace*. His is, however, a historian's book, and the book of a political scientist more interested in the forces which influence the level and allocation of food aid flows than in those which affect its impact on development.

Wallerstein's book is divided into four parts and 12 chapters. Part I presents a fairly straightforward, factual summary of the origin and subsequent evolution of US food aid policy and Part II does the same for other bilateral donors and the UN system. An important insight growing out of the historical material on US policy is the perception that in periods of crop surplus domestic agricultural policy concerns have dominated the political process, whereas in periods of shortage foreign policy concerns have prevailed. This means that if agricultural experts are correctly reading future trends, and world grain markets do not return to a situation of chronic surplus, foreign policy considerations will be even more important than before in fixing the level and direction of US food aid flows.

Economic development in friendly Third World countries has been and will continue to be an important objective of American foreign policy, and food aid will probably continue to be an important instrument for attempting to achieve this aim among others. In Part III, Wallerstein presents case material for each major recipient country, showing how the food aid programme had served one or more of the following foreign policy purposes: to give economic support to Cold War allies; to secure military bases; to exercise diplomatic leverage in tense international hotspots; to maintain friendly relations with UN voting allies. He also discusses the

Seeds of Plenty, Seeds of Want: Social and Economic Implications of the Green Revolution. By Andrew Pearse. Pp.262. ISBN 0-19-877150-9. (Clarendon/Oxford University Press: 1980.) £7.50, \$22.50. *Food for War — Food for Peace: United States Food Aid in a Global Context*. By Mitchel B. Wallerstein. Pp.312. ISBN 0-262-23106-9. (MIT Press: 1980.) \$30, £18.60.

interaction between the use of food aid to foster US trade interests and the withholding of it as a factor in US economic and commercial relations with Third World countries. An important aspect of Wallerstein's treatment is his focus on the individual concerns and policy interests of each of the four Presidents who have exercised control over US food aid policy since the shift toward its greater use for foreign policy purposes in the early 1960s. The effect of personality and temperament was clearly important, and this suggests that while the categories Wallerstein uses may be appropriate for the specific circumstances of the past, they will not necessarily fit the future so neatly.

The final part of the book treats issues relating to the role of multilateral versus bilateral food aid in the context of international politics. There are some cogent reasons for both donors and recipients to prefer that some portion of the overall flow of food aid be channelled through multilateral assistance agencies. However, both international politics and the sluggishness of the US bureaucracy constrain the amount which can be effectively administered in this way. An alternative for the future which Wallerstein suggests, and which is already a glimmering in the eyes of some of the more forward-looking participants in the food aid debate, would be for the International Wheat Council to play an increased role in the coordination of food aid flows, and for donors to come together more frequently through aid consortia for agricultural development. This mechanism would encourage the working out of the respective roles of food aid and other foreign assistance in complementary and developmentally effective fashion, without requiring donors to give up the foreign policy benefits they stand to gain from maintaining bilateral food aid

programmes.

Like many good historians, Wallerstein uses his material to weave a story which is full of complexity and detail, bringing into focus the pattern of past events without prejudging the future. In a sense, Pearse's book is also historical, although he tries so hard to be true to the results of each piece of research which he reports that the story he has to tell suffers somewhat. While it is not an easy book to read, *Seeds of Plenty, Seeds of Want* does have some important points to make. It is a book about smallness versus largeness, about cultivation for a livelihood versus cultivation for a profit, about the old way of doing things versus the new. It raises issues and suggests problems arising out of Green Revolution strategies of agricultural development which concern aid donors and recipients alike.

Pearse's volume grew out of a joint undertaking of the UN Research Institute for Social Development/UN Development Programme (UNRISD/UNDP) which he directed, the brief of which was to examine the socio-economic effects of the Green Revolution. The work began in 1970 and was conducted by a number of researchers, each working on an individual case study or particular aspect of the subject. Much of the research was completed by 1974 and many of the case-study results have subsequently been published in their own right. Nevertheless, the project directors rightly felt that an analytical and interpretative summary of the main conclusions and their policy implications would be in order, and Pearse's book is the result.

Since the volume is a summary of previously published reports of individual research efforts, some of them dating back several years, its analysis and conclusions do not always reflect the latest results available; nor is reference made to similar kinds of studies done by workers who were not part of the UNRISD study. This is unfortunate, since some later work modifies and amplifies the findings of the early UNRISD pieces, and the book would have had more general value if it had set the UNRISD material in the context of the larger body of work which addresses questions relating to the impact of Green Revolution technology and means of improving its effectiveness as a tool for agricultural development. Even among those who originally contributed, there is

growing recognition that, as the full effects of new technologies continue to work themselves out, dynamic changes are beginning to occur which will provide new economic opportunities and more widely distributed income growth than previously thought.

This difficulty does not detract from the main thrust of the argument, however. Pearse states frankly that his objective, and that of his collaborators in the UNRISD studies, is "to strengthen those forces likely to press for non-elitist development" (p.219), and the concluding part is particularly successful in creating the basis for doing this. The first two parts attempt to draw generalizations from the specific case studies that were part of the UNRISD project. Here Pearse meets with less success — the material is organized somewhat arbitrarily around a number of themes commonly associated with Green Revolution impact analysis. However, some of the studies are quite situation specific and of doubtful general relevance.

In Part III the critical issues are summarized in Pearse's own words, and here both the nature of specific difficulties and the lines of possible policy change are spelled out with great clarity. It is not simply that the rich are getting richer and the poor poorer. More importantly, the economics of agricultural development are such that the most effective techniques for increasing agricultural production result in a reduction in welfare for small farmers where there is land scarcity and labour surplus. Pearse does not oversimplify and make the mistake of asserting that a return to subsistence agriculture can solve the food problem; he recognizes that production growth cannot occur without commercialization of farming. But he points out that under the early Green Revolution strategy, the number of sub-livelihood households increased, with no comparable increase in alternative earning opportunities for the growing number of landless labourers.

Thus he recommends the different strategy of concentrating on the poor but aspiring farmer as initial adopter of new agricultural technologies, and focuses on questions of appropriateness of technologies, institutions, and price and planning policies for implementing such a strategy. Out of the questioning of the old approach and the recognition of previous failures has arisen a new "food systems" approach to planning which some countries are already beginning to pursue and which is being supported by further research at UNRISD. With its focus on the distributional as well as the yield-increasing results of agricultural development strategy, the new approach promises to open up different avenues for seeing and doing which give hope that one day a book such as this will no longer be necessary. □

Barbara Huddleston is a Research Fellow at the International Food Policy Research Institute, Washington DC.

Astronomical, in content and in effort

Michael Salt and Joseph Needham

Chung-Kuo Ta Pai K'o Ch'üan Shu (The Greater Encyclopaedia of China). Vol.1 *T'ien Wên Hsüeh* (Astronomy). Pp.650. (Greater Encyclopaedia of China Publishing House, 1-A Wai Guan Dong Jie, Beijing, 27 (Peking), China: 1981.) \$30.

"Your writing's grown weedier and weedier:
Produce more, or we've no further need o' ya",
His publishers warned.
Their advice wasn't scorned:
He's at work on an encyclopaedia.

At the present time we seem to be flooded with encyclopaedic compilations. Last year, there was the new *Britannica*, and the 16-volume *Dictionary of Scientific Biography*. This year has seen the new and revised edition of *Grove*. In popular astronomy alone, half-a-dozen different compendia are available.

But China is different, for there has been no grand encyclopaedia produced in that great culture in modern times at all. The present book is the first of 80 volumes that will cover archaeology, law and jurisprudence, history, literature, philosophy, religion, economics and many other subjects. It deals not only with contemporary astronomy but also with its history. Astronomy is one of the few sciences for which it is possible to write a single, continuous, historical account reaching from antiquity to the present day, as was done by R. Wolf in his *Handbuch der Astronomie; ihrer Geschichte und Litteratur* of 1890. Now Chang Yü-Chê and his 20 colleagues have accomplished the same.

Physically, this first volume is of excellent quality. The binding (of our hardback copy, at least) has survived the destructive testing to which so many Chinese books succumb in transit; the paper is thin but substantial; and the print is clean and easy to read. The numerous inset photographs and diagrams, too, are of admirable quality. So are the beautiful coloured plates, clustered together at intervals throughout the book. Several hours of serious reading have revealed only two misprints, though in a volume of one and a half million (Chinese) characters, there must presumably be others. But our first and enduring impression is that proof-readers and publishers have done their utmost to ensure an accurate text, handsomely presented.

On the whole, the book is easy to use. An English index to entries rescues those whose technical Chinese is insufficient to give them accurate bearings in the vastness of the field. A glossary of proper names, in Roman and Cyrillic scripts, promises help with the transliterated scientists who feature in the text. This promise is not per-

haps fully carried out, however, for several names (e.g. de Broglie, Lorentz, Narlikar and Planck) are not given in the glossary. This list is really a conversion table from the Western forms to the Chinese forms, and one has to be forewarned that in the main index the alphabetical order follows the Chinese form of the name, not the original Western one. It is a pleasure to see that due credit is given to the great scientific men such as Chang Hêng, Tsu Ch'ung-Chih, I-Hsing and Kuo Shou-Ching, who figure in the history of astronomy; though we cannot say we like the practice of giving entirely imaginary portraits of them (as in Pl.5), recalling the Renaissance busts of Aristotle and Galen with which the older historians of biology and medicine used to ornament their books. There are also plenty of plates and text-figures showing ancient Chinese instruments, and pages of old Chinese books, all well explained, but they do not in any way impede the exposition of the most modern theories and the most recent knowledge. Finally, star-maps identify the traditional Chinese constellations in terms of those used scientifically today.

There are minor inconveniences. One of these arises from the purely alphabetical arrangement of the entries, which follows the sounds of the Pin-yin romanization. Another concerns the diagrams. In a work arranged by subject, a few well-placed and well-chosen diagrams can be made to serve an entire section, but here diagrams are sometimes lacking in places where they would be helpful. For instance, in the article on relativistic cosmology (p.444), there is no diagram to illustrate the different values of k in the three models of the Universe; but a suitable graph is given in the entry on the age of the Universe (p.519). Unfortunately, though, there is no cross-reference.

Cross-references to entries within the volume are achieved by a simple convention; a different Chinese type-face in the text. In practice, though, this is not rigorously applied. For instance, the entry on the "problem of the advance of Mercury's perihelion" (p.333) mentions Le Verrier and Newcomb, but fails to indicate that each of these men is himself the subject of an entry (pp.204, 244).

The matter of entries leads to two more important questions. First: in an encyclopaedia of astronomy, how much physics, say, or mathematics, should be included? Second: how much background knowledge, and what level of understanding, should the compiler expect of his readers? These questions face every author. But for the encyclopaedist, they are of paramount importance.

As regards coverage, we are not competent to judge whether or not all the growing points of this prolific subject are

adequately covered. We were, however, slightly disappointed to find no separate articles on quantum theory, on the special or general theories of relativity, or on tensor analysis. One wonders why such topics as these are not separately expounded when the scintillation counter (p.288), photoneutrino processes (p.96) and metric (p.64) each feature individually? Future volumes will no doubt supply the missing information, but for the present our thirst for understanding is not slaked.

The second question is also difficult to answer on the basis of the present volume. In general, the articles are well written. Some, indeed, are admirable, both in exposition and content. Others are less good, either failing to link the subject in hand with cognate subjects, and with the general principles involved; or attempting to cover basic theory, the history of the subject and its interrelations, all in one and the same exposé. Moreover, not all articles start from the same intellectual baseline. In one, for instance, π and G are glossed, but in another, a partial differential equation is

given, without preliminaries or explanation. This case suggests that this volume is intended primarily for scholars and research workers. But if so, why do most of the articles lack bibliographies? And why are such references as are given always to books rather than to research papers?

But these are early days. We must not focus disproportionately on comparatively minor shortcomings. Subsequent volumes will surely make further improvements even on the high standard that this first volume has set. This work is of a very high standard, quite dazzling in production. Contributors and editors alike should rightly feel proud of all that they have achieved within three years. We look forward eagerly to the remainder of what promises to be a magnificent compilation.

Michael Salt and Joseph Needham are at the East Asian History of Science Library, Cambridge, UK. The most recent of Dr Needham's volumes in the series Science and Civilisation in China was published late last year.

Prehistoric Sahara of green and plenty

Martin Williams

Prehistory of the Eastern Sahara. By F. Wendorf and R. Schild. Pp.414. ISBN 0-12-743960-9. (Academic: 1980.) \$65, £24.20.

DEFUNCT lakes and rivers, plant and animal fossils, Neolithic villages and rock paintings testify to when the Sahara was, in certain parts at least, a green and pleasant land. Nowhere is the contrast between present wasteland and prehistoric plenty more stark than in the Western Desert of Egypt. For six seasons, from 1972 to 1977, a multidisciplinary team from Egypt, Poland and the United States laboured to reconstruct the life-style, economy and habitat of the Palaeolithic and Neolithic peoples who once inhabited the now-arid wastes between Libya and the Nile.

The book opens with a short, clear account of field and analytical methods. Four profusely illustrated chapters detail the archaeological finds and stratigraphy in each of the localities studied, after which the geographically and chronologically discrete strands are woven into two rich tapestries depicting the former environments and human exploitation of the Western Desert. A lucid, thought-provoking discussion of the origin of food production in northern Africa concludes the main section of the book. Ten specialist appendices follow, including evaluations of Quaternary deposits (marred by the unreadable Figure A1.1), radiocarbon dating problems, artefact and pottery analyses, the fossil record, and a scanning electron microscope study of cereal grains.

What conclusions emerge from these new data? The most interesting, best documented evidence relates to the Terminal Palaeolithic and Neolithic sites near Gebel Nabta, 100 km west of Abu Simbel. At Nabta playa three Holocene lake phases date to ≥ 9000 , ≥ 8600 and 7000–5800 BP. Terminal Palaeolithic sites persist until the end of the first moist phase; Neolithic sites appear at the close of the ensuing dry phase. Early cattle, sheep and goat remains date respectively to 9300, 8100 and 7000 BP. Attempts by the early desert dwellers to adapt to an increasingly harsh Holocene climate culminated in domestication of animals and plants over a thousand years before food production became widespread in the adjacent Nile valley.

The final exodus of Neolithic herders from the dying grasslands and shrinking lakes of the Saharan interior must have had repercussions beyond the confines of the desert. In this context, the recent migrations of starving herds and emaciated nomads during the latest sahel drought reflect the tragic but inexorable late Holocene desiccation of our greatest desert.

A fine testimonial to the dedication of the archaeologists and to the organizational skills of the two authors, this clearly written, well-produced volume deserves to be widely read by prehistorians and students of the Quaternary. □

Martin Williams is Associate Professor in Earth Sciences and Director of the Quaternary Research Unit at Macquarie University, Australia.

Blossoming Balkans

P.J. Grubb

Flowers of Greece and the Balkans: A Field Guide. By Oleg Polunin. Pp.592 + plates. ISBN 0-19-217626-9. (Oxford University Press: 1980.) £40, \$125.

THE latest in Oleg Polunin's series of illustrated guides to European plants is a most attractive introduction to the Balkan flora, which is the richest in Europe, amounting to more than 5000 species. Over 500 of these are illustrated in colour photographs and some 350 others in black-and-white drawings. The plan of the book is essentially the same as that adopted in the earlier *Flowers of South-west Europe* (OUP, 1973), produced jointly with B.E. Smythies. There is a general account of the climate, geology, flora and vegetation, an introduction to each of the "plant hunting regions", and then brief diagnoses of about 2000 species (320 pp.), followed by indexes of names and places, and an exhaustive bibliography. A new departure is the index of popular names used in Yugoslavia, Bulgaria and Greece. There are 22 absolutely superb colour plates of representative landscapes, but my copy lacks numbers 1 to 4 — an unreasonable fault in a book costing £40!

The author conveys his "feel" for the country and his enthusiasm for the plants very well. His accounts of particular regions have been vetted by local botanists, and those of the several areas I have visited myself ring true. The drawings are excellent, and the photographs constitute the usual mixture, some being of little value for critical identification but others are perfect. Whether the accounts of genera that are nowhere illustrated will be of any use to those who have not seen the plants concerned elsewhere, I doubt.

The weakest part of the book is the ecological account, particularly the section on forest and scrub communities, and it would have been better to have followed more closely the scheme of Horvat, Glavac and Ellenberg, whose map of vegetation-types has been used as the basis for the simplified map lining the covers. More precise thumb-nail sketches of the ecology of each species described would also have been possible. The worst fault is the emphasis on 1920s' ideas of "migration routes" of plants, and the failure even to mention the appreciation built up in the 1960s and 1970s of the dramatic changes that occurred in Mediterranean communities during the Ice Ages, when steppe seems to have replaced much of the forest.

Despite its minor blemishes, this book will be an invaluable guide in the Balkans for those who already have a good knowledge of northern and central European plants. □

P.J. Grubb is a University Lecturer in Botany at the University of Cambridge.

Life of an archaeologist extraordinary

Peter Warren

The Find of a Lifetime: Sir Arthur Evans and the Discovery of Knossos. By Sylvia L. Horwitz. Pp.278. ISBN 0-670-13575-5. (Viking: 1981.) \$14.95. To be published in the UK in October by Weidenfeld & Nicolson, price £9.95.

Few subjects could be more compelling for a biographer than Sir Arthur Evans. The reasons for this are not simply external: his passionate devotion, based on deep historical study and extensive topographical research, to the freedom of the Slav peoples; his re-foundation, almost creation, of the Ashmolean Museum; his virtually personal discovery of the Minoan, Bronze Age civilization of Crete, with its complex yet graceful society; and his wealth, so generously given over to recreating the Minoans for all the world. All these stand as beacons for any biographer. But what gives the subject a yet stronger appeal is the astonishing range and quality of Evans's mind, that of a true Renaissance man, as learned and appreciative in numismatics, botany, Balkan topography and history, folklore, Celtic and Roman settlement in Britain, as in Minoan archaeology.

Yet one turns to a new biography with some trepidation, since this work was done in a most perceptive and scholarly way by Joan Evans in *Time and Chance* (Longman, Green, 1943), published within two years of Evans's death at the age of ninety. As Sir Arthur's half-sister, Joan Evans had access to — indeed was an intimate part of — family history, which made her book a masterpiece. Since that time, too, the archaeologist W.A. McDonald has included a lengthy chapter on our subject in his excellent survey, *Progress into the Past*. How, then, does Sylvia Horwitz fare? The answer is that hers is very good and sympathetic account. While its lighter tone and air of a rapid read mean it in no sense replaces *Time and Chance*, its qualities are clear. It is accurate, a few smallish errors apart; it is lively and enjoyable, clear though brief on archaeological and Balkan political issues; it does not disguise Evans's autocratic and imperious attitudes, nor his generosity. The most perceptive sections are those examining the effects upon Evans of the death of his mother when he was six, of the sister closest to him and of his wife Margaret, daughter of the Liberal and historian Edward Freeman, after only 15 years of marriage. The author's suggestions are convincing, that these losses made Evans deeply withdrawn and much more reticent about personal feelings than was usual even for a high Victorian, while at the same time, in pursuing his scholarly and political interests, and with children, he was extrovert, generous and always vigorously engaged.

The book makes good use of the 40 years

since Evans's death and the publication of *Time and Chance* by describing how his theories and interpretations have fared after two more generations of archaeologists and Cretan discoveries. With the exception of his views on Mycenae, the huge edifice of thought he erected, on religion, art, architecture and Cretan prehistory, remains the firm basis of contemporary Minoan understanding. While reaction to his Minoan colonization of the Greek mainland was natural, given the rich vein of Mycenaean discoveries explored by A.J.B. Wace and a distinguished line of Greek archaeologists, the extent of Cretan artistic and economic influence throughout the south Aegean and southern Peloponnese now appears to be much greater than critics of Evans's views could then allow. His dating of the destruction of Knossos as a palace, at around 1400 BC, and all that that means for Aegean prehistory, came under detailed attack from Professor L.R. Palmer some years ago; but a large majority of archaeologists has confirmed the essential correctness of Evans's and his assistant Mackenzie's views. Thermoluminescence dating may one day provide a close date for the burning of the Linear B tablets, at the heart of that controversy, in the fire which consumed the palace.

Unlike recent studies on Schliemann, Sylvia Horwitz's account is thoroughly

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Sir Arthur Evans — a true Renaissance man.

sympathetic. She admires the vast nervous energy, the astonishing yet repeated flair for understanding an object or its function from exiguous evidence, often confirmed years later by new discoveries. She found no need for a severe critique of this eminent Victorian. Her account can be warmly recommended, for itself and as a preliminary to *Time and Chance*. We should recall only that as much is to be found of political reporting from remote Bosnian fastnesses or richly flowered Ragusan gardens as of the splendours of the House of Minos. □

Peter Warren is Professor of Ancient History and Classical Archaeology at the University of Bristol, and currently Director of Excavations at Knossos.

Ordered amino acids brought to book

Linda Fothergill

Handbook of Protein Sequence Analysis: A Compilation of Amino Acid Sequences of Proteins with an Introduction to the Methodology. 2nd Edn. By L.R. Croft. Pp.628. ISBN 0-471-27703-7. (Wiley: 1981.) £38, \$104.50.

THIS second edition of Croft's compilation of amino acid sequences of proteins is strikingly different from its predecessor, both in format and content. Disappointingly, it is no longer loose-leaf, thereby losing the useful flexibility of the first edition. It is obviously no longer possible to slip in annual supplements — a most attractive feature in a field such as protein (and nucleic acid) sequence determination that expands at such a dauntingly rapid rate. Sadly, also the useful transparent overlay to show residue numbers is no longer supplied, although the extraordinary practice of designating N-terminal acetyl groups as residue one has been perpetuated in the second edition.

A new feature of this second edition is that approximately one-third is devoted to protein sequence determination

methodology. Much of this section is admirably full of practical detail — something unfortunately often lacking in primary publications. There are some gaps however (HPLC separation of peptides is not mentioned for example), and a more extensive, separate publication would seem more satisfactory. Moreover, I suspect that a great many of those consulting a handbook of sequences are not particularly interested in the methodology.

A comparison of this handbook with Dayhoff's *Atlas of Protein Sequence and Structure* (Supplement 3 to Vol.5, 1978) shows that Croft's book is more up-to-date, but that it includes fewer homologous sequences from related species, no fragments and less extensive notes on the proteins.

Croft's book will be indispensable in providing the most up-to-date handbook of protein sequences currently available; but please, for the next edition, return to the loose-leaf format with annual supplements! □

Linda Fothergill is a Research Officer in the Department of Biochemistry, University of Aberdeen.